

School of Pharmacy

**Molecular Mechanisms Underlying Aberrant Expression of the
Connective Tissue Growth Factor in Paediatric Pre-B Cell Acute
Lymphoblastic Leukaemia**

Mathew D. Welch

This thesis is presented for the Degree of

Doctor of Philosophy

of

Curtin University

June 2011

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Mathew D. Welch

Date: 10th Jan 2012

Acknowledgements

I would like to acknowledge a number of people without whom the work presented in this thesis would not have been possible.

To my supervisors Ursula Kees, Wayne Greene and Brett Dix, thankyou for your encouragement and guidance throughout this journey. Your continued passion for medical research and countless hours spent helping me throughout the project really made a huge difference.

To my mentor Deborah Strickland, thanks for your unwavering encouragement and advice. Thanks to all the past and present members of the division of Leukaemia and Cancer Research at the Telethon Institute for Child Health Research, particularly Laura Genovesi and Amy Samuels for always making me laugh and picking me up when I was down.

To my family for their support, and of course to my beautiful wife Bae, for putting up with the endless weekends in the lab and the emotional roller coaster that is the life of a PhD student. Your patience, love, friendship and honesty make you the most amazing person in my life and there is absolutely no way I could have done this without your support.

Abstract

Acute lymphoblastic leukaemia (ALL) is the most common cancer diagnosed in children aged 1-14 years. There have been vast improvements in clinical outcomes for children diagnosed with ALL with cure rates of up to 90% achievable for some forms of the disease. Despite these successes, some patients still relapse and the prognosis for these individuals is poor, thus there is still a great deal to be learned about the complex biology underlying ALL. Connective tissue growth factor (*CTGF/CCN2*) is a novel candidate gene in precursor B-cell (pre-B) ALL, and is aberrantly expressed in a high proportion (around 75%) of cases. While the CTGF protein has no known role in lymphocyte biology or haemopoiesis, *CTGF* gene expression has been associated with a poor outcome in children and adults, particularly in those patients deemed to have a high-risk of relapse. The primary aims of this study were to characterise *CTGF* expression in paediatric pre-B ALL tumours and cell lines, and to investigate what mechanisms were responsible for its deregulated expression by examining the contribution of both genetic and epigenetic factors.

Analysis of a cohort of 73 primary paediatric pre-B ALL specimens, confirmed *CTGF* was aberrantly expressed in 75% of patients at heterogeneous levels. *CTGF* expression was found to be associated with prognosis in this cohort, as there was a trend toward lower 5 year relapse free survival (RFS) in patients with *CTGF*^{pos} ALL (71% RFS in *CTGF*^{pos}, 83% RFS in *CTGF*^{low/neg}), however this did not reach statistical significance ($p=0.39$). There was no difference in overall 5 year survival (OS) between patients with *CTGF*^{pos} and *CTGF*^{low/neg} ALL. The association between *CTGF* expression and clinical features recorded at the time of biopsy was undertaken. These features included; age, gender, percentage blasts in bone marrow, peripheral haemoglobin level, spleen or lymph node involvement, or whether the sample was obtained at diagnosis or relapse. Patients with enlarged lymph nodes displayed a lower mean *CTGF* expression compared to those patients with no lymph node involvement. A similar pattern was observed with patients exhibiting enlargement of the spleen, however this did not reach statistical significance. No other clinical features were associated with *CTGF* expression.

Analysis of global gene expression patterns in three independent paediatric pre-B ALL cohorts (PMH; n=73, Ross; n=118, Kang; n=207), identified five genes that were highly correlated with *CTGF* expression; *SOCS2*, *MEF2C*, *ADD3*, *GSN* and *DPYSL2*. *In silico* analysis of the 5' flanking sequences of these genes, as well as *CTGF* identified predicted binding sites for the two Ikaros family proteins IKAROS and HELIOS, with at least one HELIOS site present in the 5' flanking sequence of all correlated genes. This suggested a possible role for Ikaros family proteins in modulating *CTGF* gene expression. Subsequent analysis of a microarray cohort recently characterised for *IKAROS* gene deletions revealed that mutations or deletions within the *IKAROS* coding region were significantly associated with higher mean *CTGF* gene expression, implying that a loss of IKAROS function may promote activation of the *CTGF* locus. This is the first evidence implicating the Ikaros family of lymphoid transcriptional regulators in deregulation of the *CTGF* locus.

Northern blotting of RNA from B-lineage ALL cell lines uncovered evidence of alternative splicing of *CTGF* mRNA. Non-canonical transcripts of approximately 1.3kb and 1.6 kb were hybridised by a *CTGF*-specific probe in extracts from all *CTGF*^{pos} cell lines. Sequencing of cDNA fragments as well as 5' and 3' RACE products from a cell line with the highest level of *CTGF* expression revealed a number of *CTGF* transcripts exhibiting internal deletions of exons 2 and 3, as well as truncation of exons 1 and 4. 3' RACE also identified a 1.3kb transcript that was devoid of 3' UTR regulatory elements as a result of premature polyadenylation. These findings represent the first direct evidence of alternative splicing of *CTGF* pre-mRNA in any tissue type and further investigation is warranted to fully characterise these tumour-associated transcripts and their protein coding potential.

Structural changes within the genome are common in leukaemia and deletions affecting the long arm of chromosome 6 occur in around 30% of cases of pre-B ALL. To investigate whether deregulation of *CTGF* expression has a genetic basis, the *CTGF* locus at 6q23.1 was investigated for structural alterations or mutations. Southern blotting performed in seven B-lineage ALL cell lines (4 *CTGF*^{pos}, 3 *CTGF*^{neg}) confirmed that the

CTGF locus was not cytogenetically rearranged. Furthermore, analysis of *CTGF* copy number by qPCR in primary paediatric pre-B ALL specimens (n=17) and B-lineage ALL cell lines (n=7) confirmed that gene amplification could not account for *CTGF* overexpression. Sequencing of the *CTGF* promoter and 3' UTR in three B-lineage ALL cell lines confirmed that there were no mutations affecting these important regulatory regions, although one cell line harboured the rs6918698 -739 C>G SNP, which is predicted to disrupt SP3-mediated repression of *CTGF* gene expression.

Epigenetic regulation of gene expression is frequently altered in neoplasia. The *CTGF* locus contains a CpG island, and methylation of this region was found to be inversely correlated with *CTGF* gene expression in B-lineage ALL cell lines, as assessed by both methylation-specific PCR and by bisulfite sequencing. Bisulfite sequencing of primary tumour specimens revealed that hypomethylation of the *CTGF* locus was a widespread feature of pre-B ALL. By contrast, analysis of primary T-ALL specimens demonstrated extensive methylation at the *CTGF* locus, indicating that *CTGF* may be permissive for expression specifically in pre-B ALL.

This study has highlighted several novel aspects of *CTGF* expression in pre-B ALL, including a potential role for Ikaros family proteins in regulating the *CTGF* locus, and the existence of *CTGF* mRNA transcripts generated through alternative pre-mRNA splicing. Investigation of mechanisms promoting *CTGF* gene expression in pre-B ALL revealed that the rs6918698 C>G SNP was present in one pre-B ALL cell line. Hypomethylation of the *CTGF* locus was a notable feature of primary pre-B ALL specimens and was in contrast to the hypermethylation observed in 2 T-ALL specimens and CD34^{pos} bone marrow cells. These findings will direct future research to elucidate the complex mechanisms regulating *CTGF* expression in pre-B ALL. It is anticipated that illuminating the role of *CTGF* in the pathogenesis of ALL may result in significantly improved patient outcomes through the development of targeted and less toxic therapies, and through improved risk-based stratification of patients to ensure those at a high risk of relapse are directed toward an appropriate level of therapeutic intervention.

Abbreviations

The following abbreviations are used throughout this thesis:

3'UTR	3' untranslated region
5'UTR	5' untranslated region
ALL	Acute lymphoblastic leukaemia
AU	Above umbilicus
Aza	5-aza-2'-deoxycytidine
BAC	Bacterial artificial chromosome
BCE-1	Basal control element
bHLH	Basic helix-loop-helix
BMSC	Bone marrow stromal cells
BU	Below umbilicus
CAESAR	Cis-acting element of structural-anchored repression
CAF	Cancer-associated fibroblast
CCG	Children's Cancer Group
CCR	Complete clinical remission
CD	Cluster of differentiation
cDNA	Complimentary DNA
CGH	Comparative genome hybridization
ChIP	Chromatin immunoprecipitation
CLP	Common lymphoid progenitors
CM	Conditioned cell culture medium
CMP	Common myeloid progenitors
CAN	Copy number alteration
CNS	Central nervous system
COG	Children's Oncology Group
CT	Carboxy-terminus
C_t	Cycle threshold
DN	Dominant negative
DNA	Deoxyribonucleic acid

EFS	Event free survival
EST	Expressed sequence tag
FACS	Fluorescence activated cell sorting
FCS	Fetal calf serum
GSP	Gene specific primer
HDAC	Histone deacetylase
HPC	Haemopoietic progenitor cell
HRE	Hypoxia inducible factor 1 response element
HSC	Haemopoietic stem cell
IG	Immunoglobulin
IGFBP	Insulin-like growth factor binding proteins
m⁵C	5' methylcytosine
MEM	Meme-enriched motif
MMP	Matrix metalloprotease
MPP	Multipotent progenitor cell
MRD	Minimal residual disease
NTC	Non-template control
OH-m⁵C	5' hydroxy-methylcytosine
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
POG	Pediatric Oncology Group
pre-B ALL	Precursor B-cell acute lymphoblastic leukaemia
qPCR	Quantitative polymerase chain reaction
qRT-PCR	Quantitative real-time polymerase chain reaction
RACE	Rapid amplification of cDNA ends
rCTGF	Recombinant CTGF
RFD	Reticulin fibre density
RFS	Relapse free survival
RNA	Ribonucleic acid
RMA	Robust multi-array analysis
RT	Reverse transcriptase

RT-PCR	reverse-transcriptase polymerase chain reaction
SBE	Smad-binding element
SDS	Sodium dodecyl sulphate
SMP	Skim milk powder
SNP	Single nucleotide polymorphism
SP	Signal peptide
T-ALL	T-cell acute lymphoblastic leukaemia
TGFB1	Transforming growth factor beta 1
TBRE	Transforming growth factor beta 1 response element
TCR	T-cell receptor
TRENDIC	Transcriptional enhancer dominant in chondrocytes
TSA	Trichostatin-A
TSG	Tumour suppressor gene
TSP	Thrombospondin type-1
TSS	Transcription start site
U	Unit
UKCCSG	United Kingdom Children's Cancer Study Group
VWC	Von Willebrand type-C domain
WBC	White blood cell

List of figures and tables

Figure 1.1	Model of human B-cell development	5
Table 1.1	Immunologic classification of B-lineage ALL	10
Table 1.2	Clinical and biological features for initial patient stratification	11
Table 1.3	Common cytogenetic abnormalities in pre-B ALL	13
Figure 1.2	Modular structure of CCN proteins	23
Figure 1.3	Transcription factor binding sites at the <i>CTGF</i> promoter	30
Table A2.1	List of suppliers.....	60
Table A2.2	Primer and oligonucleotide sequences.....	61
Table 3.1	Clinical and biological features of pre-B ALL patients in the PMH cohort	66
Table 3.2	Characteristics of pre-B ALL specimens used for validation of <i>CTGF</i> expression	67
Figure 3.1	Analysis of <i>CTGF</i> expression by microarray and qRT-PCR	69
Figure 3.2	<i>CTGF</i> correlation analysis in three independent cohorts	72
Figure 3.3	MEME enriched motif 1 (MEM1)	74
Figure 3.4	MEME enriched motif 2 (MEM2)	75
Figure 3.5	DNA binding matrices for the Ikaros family members 1, 2 and 3	76
Figure 3.6	Helios (IKZF2) is predicted to bind the promoters of <i>CTGF</i> and correlated genes	78
Figure 3.7	Expression of Ikaros family members in the PMH cohort	79
Figure 3.8	<i>IKAROS</i> alterations are associated with aberrant <i>CTGF</i> expression in high-risk pre-B ALL.....	81
Figure 3.9	<i>CTGF</i> gene expression in the PMH cohort.....	82
Figure 3.10	Comparison of <i>CTGF</i> expression versus clinical features present at diagnosis in the PMH cohort	84
Figure 3.11	Association between <i>CTGF</i> expression and patient outcome.	85
Table 4.1	Features of B-lineage ALL cell lines	95
Figure 4.1	<i>CTGF</i> gene expression in ALL cell lines.....	96
Figure 4.2	Detection of CTGF protein by immunoblotting	99

Figure 1.3	Transcription factor binding sites at the <i>CTGF</i> promoter	101
Figure 4.4	Analysis of the rs6918698 C>G SNP in PER-278	102
Figure 4.5	miRNA binding sites in the <i>CTGF</i> 3' UTR	103
Table 4.2	Documented aberrations involving 6q23 in ALL	104
Figure 4.6	Expression of genes neighbouring the <i>CTGF</i> coding region	105
Figure 4.7	Localisation of hybridisation probe and restriction sites used for Southern blotting analysis of the <i>CTGF</i> locus	107
Figure 4.8	Southern blot analysis of the <i>CTGF</i> locus	108
Figure 4.9	<i>CTGF</i> copy number analysis	110
Figure 4.10	Raw gene copy number data obtained from ALL cell lines	111
Table 5.1	<i>CTGF</i> mRNA expression in B-lineage ALL cell lines	119
Figure 5.1	Expression of <i>CTGF</i> mRNA transcripts in the PER-377 cell line....	120
Figure 5.2	Characterisation of <i>CTGF</i> mRNA transcripts in ALL cell lines	122
Figure 5.3	Structure of <i>CTGF</i> Variant 1	124
Figure 5.4	Structure of <i>CTGF</i> Variants 2 and 3	126
Figure 5.5	<i>CTGF</i> transcript variants may retain protein coding potential.....	128
Figure 5.6	Amplification of <i>CTGF</i> variants by 5' and 3' RACE.....	130
Figure 5.7	Re-amplification of individual <i>CTGF</i> RACE products by band-stab PCR.....	132
Figure 5.8	Southern blotting of 5' RACE PCR products.....	133
Figure 5.9	Characterisation of PER-377 <i>CTGF</i> 5' RACE products	136
Figure 5.10	Characterisation of PER-377 <i>CTGF</i> 3' RACE products	139
Figure 5.11	Sequence analysis of 3' RACE clone, 3_5.....	140
Figure 5.12	Summary of novel <i>CTGF</i> transcripts detected in pre-B ALL cells..	143
Figure 6.1	CpG dinucleotide distribution within the <i>CTGF</i> locus	153
Figure 6.2	Nucleotide sequence of the <i>CTGF</i> CpG island.....	154
Figure 6.3	Methylation-specific PCR analysis of ALL cell lines	156
Figure 6.4	Methylation status determined by bisulfite sequencing of the <i>CTGF</i> locus correlates with <i>CTGF</i> expression in B-lineage ALL cell lines	159
Figure 6.5	msPCR analysis of primary pre-B ALL specimens.....	160

Figure 6.6	<i>CTGF</i> mRNA expression in primary pre-B ALL specimens used for bisulfite sequencing	161
Figure 6.7	Bisulfite sequencing of <i>CTGF</i>^{high} primary pre-B ALL specimens ...	162
Figure 6.8	Bisulfite sequencing of <i>CTGF</i>^{low/neg} primary pre-B ALL specimens	164
Figure 6.9	Bisulfite sequencing of primary T-ALL specimens and normal CD34 positive bone marrow cells.....	166
Figure 6.10	The effect of Aza and TSA on <i>CTGF</i> expression in B-lineage ALL cell lines PER-377 and PER-145	168
Figure 7.1	Proposed model of the growth-promoting effects of <i>CTGF</i> in the bone marrow microenvironment.....	178

Table of Contents

<i>Title Page</i>	i
<i>Declaration</i>	ii
<i>Acknowledgements</i>	iii
<i>Abstract</i>	iv
<i>Abbreviations</i>	vii
<i>List of figures and tables</i>	x

Chapter 1: General Introduction

1.1 Childhood cancer	1
1.2 Acute lymphoblastic leukaemia	1
1.2.1 Origins: Normal haemopoiesis	2
1.2.1.1 Haemopoietic stem cell niche	3
1.2.1.2 Lymphoid specification	4
1.2.1.3 B-cell development	4
1.2.2 Origins: Aetiology and incidence	6
1.2.3 ALL therapy	6
1.2.3.1 Induction therapy	7
1.2.3.2 Consolidation and re-induction	8
1.2.3.3 Maintenance therapy	9
1.3 B-lineage acute lymphoblastic leukaemia	9
1.3.1 Prognosis and risk stratification	10
1.3.2 Molecular types and cytogenetics	11
1.3.2.1 <i>TEL-AML1</i>	12
1.3.2.2 <i>TCF3-PBX1</i>	14
1.3.2.3 <i>BCR-ABL1</i>	15
1.3.2.4 <i>MLL</i> rearrangements	15
1.3.2.5 Numerical chromosomal alterations	16

1.3.2.6	Recurrent mutations in pre-B ALL	17
1.3.2.7	Classical oncogenes associated with pre-B ALL	18
1.3.2.8	Tumour suppressor genes in pre-B ALL	19
1.4	<i>CTGF</i> gene expression is dysregulated in pre-B ALL	20
1.5	Connective tissue growth factor	21
1.5.1	Origins	21
1.5.2	Protein structure	22
1.5.3	CTGF signalling	23
1.5.3.1	Insulin signalling	23
1.5.3.2	TGFB family signalling	24
1.5.3.4	Vascular endothelial growth factor signalling	24
1.5.4	CTGF cell surface interactions	25
1.5.4.1	LRP1 receptor	25
1.5.4.2	NTRK1/NGFR receptor complex	26
1.5.4.3	CTGF-associated adhesion	26
1.5.5	The <i>CTGF</i> locus	27
1.5.5.1	The <i>CTGF</i> promoter	28
1.5.5.2	Epigenetic regulation of <i>CTGF</i> expression	31
1.5.5.3	The <i>CTGF</i> 3' UTR	32
1.5.6	<i>CTGF</i> expression and cancer	32
1.5.7	<i>CTGF</i> gene expression in the bone marrow	33
1.5.8	CTGF, ALL and bone marrow fibrosis	34
1.6	Overall objectives of this study	36

Chapter 2: Materials and Methods

2.1	Specimens and cell lines	37
2.1.1	Patient specimens	37
2.1.2	Donor mononuclear cells	37
2.1.3	Mammalian tissue culture	37
2.1.3.1	Culture with 5-aza-2-deoxycytidine and Trichostatin A	38
2.2	Bioinformatics and statistics	38

2.2.1	Microarray data analysis	38
2.2.1.1	Establishing gene expression cut-offs	39
2.2.2	Promoter analysis	39
2.2.3	Statistical analyses	39
2.2.3.1	Spearman correlation analyses	39
2.2.3.2	Pearson correlation analyses	40
2.2.3.3	Graphing and statistical analysis of gene expression data	40
2.3	Molecular Techniques	40
2.3.1	Isolation of nucleic acids	41
2.3.1.1	Genomic DNA extraction	41
2.3.1.2	Total RNA extraction	41
2.3.1.3	Cytoplasmic RNA extraction	41
2.3.2	Polymerase chain reaction	42
2.3.2.1	Standard polymerase chain reaction	42
2.3.2.2	High-fidelity polymerase chain reaction	43
2.3.2.3	Agarose gel electrophoresis and gel extractions	43
2.3.2.4	First strand cDNA synthesis	44
2.3.3	Rapid amplification of cDNA ends	44
2.3.3.1	Preparation of RACE-ready RNA	44
2.3.3.2	RACE cDNA synthesis	45
2.3.3.3	Amplification of RACE clones by polymerase chain reaction	45
2.3.3.4	Band-stab polymerase chain reaction	45
2.3.4	Quantitative real-time polymerase chain reaction	46
2.3.4.1	Quantitation of <i>CTGF</i> gene expression	46
2.3.4.2	<i>CTGF</i> gene copy number analysis	46
2.3.5	Northern blotting of RNA	47
2.3.5.1	RNA electrophoresis and membrane transfer	47
2.3.5.2	Northern blot probe synthesis	47
2.3.5.3	Radiolabelling of cDNA probes	48
2.3.5.4	Northern blot probe hybridisation	48
2.3.6	Southern blotting	49

2.3.6.1	Restriction digest of genomic DNA	49
2.3.6.2	Electrophoresis and transfer of digested genomic DNA	49
2.3.6.3	Southern blot probe synthesis	50
2.3.6.4	Radiolabelling of probes	50
2.3.6.5	Southern probe hybridisation	50
2.3.6.6	Electrophoresis and transfer of RACE-PCR products	50
2.3.7	DNA cloning	51
2.3.7.1	Preparation of chemically competent cells	51
2.3.7.2	Cloning into the pGEM-T Easy vector	52
2.3.7.3	Transformation of chemically competent cells	52
2.3.7.4	Transformant screening and plasmid DNA extraction	52
2.3.7.5	Isolation of bacterial artificial chromosome DNA	53
2.3.8	DNA Sequencing	54
2.3.8.1	BigDye reaction purification and analysis	54
2.3.8.2	<i>CTGF</i> promoter and 3'UTR sequencing	55
2.3.9	Analysis of DNA methylation	55
2.3.9.1	Conversion of DNA by sodium bisulfite	55
2.3.9.2	BAC DNA methylation	55
2.3.9.3	Methylation specific PCR	56
2.3.9.4	Bisulfite Sequencing	56
2.3.10	Immunological detection of <i>CTGF</i> protein	57
2.3.10.1	Isolation of total cellular protein	57
2.3.10.2	Protein estimation	57
2.3.10.3	Sample preparation	58
2.3.10.4	Heparin-affinity enrichment of conditioned culture medium	58
2.3.10.5	Western blotting	58
2.3.10.6	Antibodies	59
2.3.10.7	Membrane stripping protocol	59

Chapter 3: Gene Expression and Clinical Correlates of *CTGF* Expression in Childhood Pre-B ALL

3.1	Introduction	63
3.2	Results	65
3.2.1	The PMH cohort of paediatric pre-B ALL	65
3.2.2	Validation of <i>CTGF</i> expression data	68
3.2.3	Gene expression profiles associated with <i>CTGF</i> expression	69
3.2.2.1	Common 5' flanking motifs in <i>CTGF</i> -correlated genes	73
3.2.2.2	Analysis of putative Ikaros family binding sites	76
3.2.2.3	Expression of Ikaros family genes in the PMH cohort	77
3.2.2.4	<i>IKAROS</i> gene lesions are associated with <i>CTGF</i> expression	79
3.2.3	Clinical features associated with <i>CTGF</i> expression	82
3.2.4	Prognostic significance of <i>CTGF</i> expression	85
3.3	Discussion	86
3.3.1	Clinical correlates of <i>CTGF</i>	87
3.3.2	<i>CTGF</i> expression is a prognostic biomarker	88
3.3.3	Gene expression correlates of <i>CTGF</i>	89
3.3.4	Ikaros family of transcriptional regulators	89
3.3.6	Conclusions	91

Chapter 4: Sequence and Structure of the *CTGF* Locus in Pre-B ALL

4.1	Introduction	92
4.2	Results	93
4.2.1	Cell line models of paediatric B-lineage ALL	93
4.2.2	<i>CTGF</i> expression in B-lineage ALL cell lines	96
4.2.2.1	<i>CTGF</i> is secreted by PER-377	97
4.2.3	Analysis of <i>CTGF</i> 5' flanking sequence	99
4.2.4	Sequencing of <i>CTGF</i> 3' UTR	102
4.2.5	Structure of the <i>CTGF</i> gene locus	103

4.2.5.1	Southern blotting of the <i>CTGF</i> locus	105
4.2.5.2	Gene copy number analysis	109
4.3	Discussion	112
4.3.1	<i>CTGF</i> promoter mutation was not detected in pre-B ALL	113
4.3.2	The <i>CTGF</i> locus is not structurally altered in pre-B ALL	114
4.3.3	Conclusions	115

Chapter 5: Characterisation of *CTGF* Transcription in B-lineage ALL

5.1	Introduction	116
5.2	Results	118
5.2.1	Northern analysis of <i>CTGF</i> mRNA in ALL cell lines	121
5.2.2	Amplification and sequencing of novel <i>CTGF</i> transcripts	123
5.2.2.1	Structural analysis of novel <i>CTGF</i> transcript variants	127
5.2.3	Amplification of <i>CTGF</i> transcripts by RACE	129
5.2.3.1	Isolation of RACE products by band-stab PCR	131
5.2.3.2	Validation of a <i>CTGF</i> probe for Southern blot hybridisation	132
5.2.3.3	Characterisation of 5' RACE products	133
5.2.3.4	Characterisation of 3' RACE products	137
5.3	Discussion	143
5.3.1	Aberrant <i>CTGF</i> transcription	144
5.3.2	Post transcriptional regulation of <i>CTGF</i> mRNA	146
5.3.3	Conclusions	148

Chapter 6: Epigenetic Mechanisms Influencing *CTGF* Expression in Pre-B ALL

6.1	Introduction	149
6.2	Results	152
6.2.1	Identification of a CpG island at the human <i>CTGF</i> genomic locus	152
6.2.2	Methylation correlates with <i>CTGF</i> expression in ALL cell lines	154
6.2.3	Bisulfite sequencing of B-lineage ALL cell lines	157

6.2.4	Hypomethylation of the <i>CTGF</i> locus in primary pre-B ALL	159
6.2.5	Bisulfite sequencing of primary pre-B ALL specimens	160
6.2.6	Bisulfite sequencing of primary T-ALL and CD34 ^{POS} specimens	165
6.2.7	Modulation of the global epigenetic state of B-lineage ALL cell lines	167
6.3	Discussion	168
6.3.1	The <i>CTGF</i> locus is hypomethylated in primary pre-B ALL	169
6.3.2	The importance of histone modifications	170
6.3.3	Conclusions	172

Chapter 7: General Discussion

7.1	Introduction	173
7.2	Study objectives	174
7.3	Critical findings	175
7.3.1	A possible role for Ikaros proteins in <i>CTGF</i> transcription	175
7.3.2	Clinical importance of <i>CTGF</i> expression in pre-B ALL	176
7.3.3	A model for <i>CTGF</i> -mediated leukaemogenesis	177
7.3.4	Genomic lesions did not contribute to <i>CTGF</i> expression	179
7.3.5	<i>CTGF</i> mRNA is subject to alternative splicing in pre-B ALL	179
7.3.6	<i>CTGF</i> mRNA may not reflect <i>CTGF</i> protein levels in ALL	180
7.3.7	The <i>CTGF</i> locus is hypomethylated in pre-B ALL	181
7.4	Future directions	182
7.4.1	Prognostic relevance of <i>CTGF</i>	182
7.4.2	Modeling the role of <i>CTGF</i> in ALL pathogenesis in-vivo	182
7.4.3	Factors interacting with the <i>CTGF</i> gene locus	183
7.5	Final conclusions	183

Chapter 8: References 185

Chapter 1

General Introduction

Chapter 1

General Introduction

1.1 Childhood cancer

In Australia, cancer is the leading cause of disease-associated deaths among children aged 1-14 years (Australian Bureau of Statistics 2006). This statistic is universal in the developed world, with childhood cancer outstripping deaths associated with other disease processes in children (Kaatsch 2010). While outcomes for some childhood cancers today have improved dramatically, there is still much work to be done to improve our understanding of underlying cancer biology. By enhancing our knowledge of the molecular mechanisms involved in oncogenesis we can continue to develop new and refine existing therapies to improve clinical outcomes.

Of the various types of cancers commonly affecting infants (< 1yr) and children (1-14 yrs), leukaemia is by far the most prevalent accounting for around 34% of new diagnoses, followed by cancers of the central nervous system at around 24% (Pieters & Carroll 2010). Leukaemia is a malignancy arising from transformation of cells of the haemopoietic system and there are two main types, lymphoblastic and myelogenous, owing to the two arms of haemopoiesis, myeloid and lymphoid which both arise from a common stem or progenitor cell. Acute lymphoblastic leukaemia (ALL) is the most common form of the disease in children accounting for 79% of all paediatric leukaemia diagnoses. This equates to 27% of all childhood cancers (Kaatsch 2010), thus, leukaemia represents a significant contribution to childhood mortality and morbidity. Improving our understanding of the molecular mechanisms of leukaemogenesis is critical to improve outcomes for young patients and their families.

1.2 Acute lymphoblastic leukaemia

As the name suggests, ALL is a swiftly progressive form of leukaemia, characterised by the clonal accumulation of rapidly proliferating, poorly differentiated lymphoid

precursors (Bourquin & Izraeli 2010). ALL disease can be divided into two broad biological categories, precursor B-cell ALL (pre-B ALL) and precursor T-cell ALL (T-ALL) (Gaynon 2005). Of these two types, pre-B ALL is more common, accounting for up to 85% of childhood ALL diagnoses (Pieters & Carroll 2010). Presenting symptoms invariably correlate with the level of tumour cell burden within the bone marrow. As normal marrow is replaced by leukaemic cells, haemopoiesis is impaired, resulting in blood lineage cytopenias and if left untreated ALL is rapidly fatal (Onciu 2009). Other common symptoms include fever, fatigue and joint or bone pain. Leukaemic cells can also reside outside of the bone marrow and common sites of extramedullary involvement include the blood, liver, spleen, lymph nodes, meninges and in T-lineage ALLs, the thymus.

1.2.1 Origins: Normal haemopoiesis

The haemopoietic system is responsible for the production of differentiated cells of all blood lineages. This includes lymphoid cells, incorporating B, T and NK cells, as well as those of the myeloid lineage, such as erythrocytes, megakaryocytes and granulocytes. All of these mature blood cells are formed from a common haemopoietic stem cell (HSC) at the apex of this developmental hierarchy (Mansson et al. 2009). These cells reside in the bone marrow space of long bones in specialised niches that are tightly regulated (Wilson & Trumpp 2006). HSCs are normally quiescent, but will respond to haemopoietic stress such as blood loss by re-entering the cell cycle and dividing to produce intermediate progenitor cells which in turn give rise to mature functional effector cells. Haemopoiesis progresses by limiting the developmental potential of precursors in a step-wise fashion until the fully mature blood cells are produced (Blom & Spits 2006). This process has been extensively characterised in mice and while surface antigens differ between species, the overall process of differentiation from HSC to mature blood cells is thought to be highly conserved (Welner et al. 2008; Mansson et al. 2009).

1.2.1.1 Haemopoietic stem cell niche

The bone marrow microenvironment is comprised of a number of different cell types, including reticular cells, fibroblasts, chondrocytes, apidoctyes, endothelial cells, osteoblasts and osteoclasts, as well as haemopoietic stem and progenitor cells (Ehninger & Trumpp 2011). Defining the role of each of these cell types within the context of haemopoiesis is still ongoing, however it is widely accepted that haemopoietic stem cells (HSCs) reside in specialised regions within the bone marrow microenvironment that govern HSC fate by providing extrinsic regulatory signals. These developmental sites, termed niches, provide extracellular cues through soluble factors and cell-cell interactions that are critical for guiding the maintenance, proliferation and differentiation of HSCs throughout adult life (Mendez-Ferrer et al. 2010; Nagasawa 2006; Sugiyama et al. 2006; Wilson et al. 2007).

The mouse haemopoietic stem cell is possibly the most-well well characterised of all mammalian stem cells. These cells are well defined by surface marker expression (lin^{neg}Scal^{hi}c-Kit⁺CD34⁻CD48⁻CD150^{hi}), and can reconstitute the entire haemopoietic system of irradiated mice (Purton & Scadden 2007; Wilson et al. 2007). Advances in intravital imaging have revealed the localisation of these HSCs within distinct regions of bone marrow including the endosteal surface (termed the endosteal niche) as well as adjacent to sinusoidal endothelium (termed the vascular niche) (Xie et al. 2009; Lo Celso et al. 2009), and these two niches are thought to differentially regulate stem cell fate. Evidence is now emerging for the central role of mesenchymal stem cells (MSCs) within the haemopoietic niche in maintaining the bone marrow HSC pool. In recent years, a subset of perivascular nestin positive MSCs have been identified that are critical for HSC maintenance (Mendez-Ferrer et al. 2010). These nestin^{pos} cells also express high levels of the HSC maintenance factors chemokine C-X-C motif ligand 12 (Cxcl12), osteopontin, angiopoietin-1, interleukin 7 and vascular cell adhesion molecule 1, and depletion of nestin^{pos} MSCs results in the mobilisation of HSCs away from the bone marrow to other organs (Mendez-Ferrer et al. 2010). Furthermore, the bone marrow-homing of transplanted HSCs into nestin^{pos} MSC-depleted recipients is significantly reduced. Similar to nestin^{pos} MSCs, a distinct subset of reticular cells also express high

levels of Cxcl12 and these cells dubbed Cxcl12 abundant reticulocytes (CAR) are found in close proximity to HSCs within the vascular niche and are critical for lymphoid specification of haematopoietic progenitor cells (Crisan et al. 2008; Omatsu et al. 2010).

1.2.1.2 Lymphoid specification

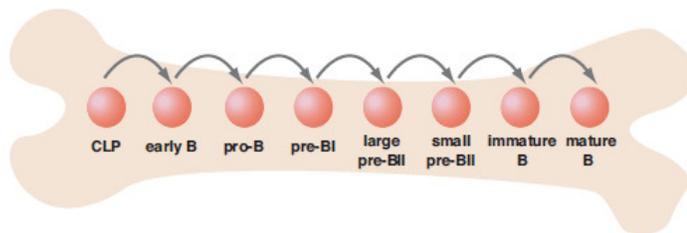
Fluorescence activated cell sorting (FACS) has allowed the prospective isolation and characterisation of cells at various stages of development with particular emphasis on their lineage potential (Mansson et al. 2009). The previously accepted model of haemopoiesis proposed a linear series of maturation steps, beginning with asymmetric division of activated HSCs giving rise to multipotent progenitor cells (MPPs), which in turn gave rise to either common myeloid progenitors (CMPs) or common lymphoid progenitors (CLP). Recent evidence has pointed to several intermediate steps in lymphoid specification prior to the restriction to CLP identity, however the precise model of transition from MPP to CLP in humans is still unclear (Corfe & Paige 2009). The earliest stage of irreversible lymphoid specification of CLPs is marked by expression of the interleukin 7 receptor (*IL7R/CD127*) (Dias et al. 2005; Parrish et al. 2009). These cells can progress to form mature B and T lymphocytes and NK cells.

1.2.1.3 B-cell development

Humoral immunity is critical to mammalian homeostasis and B cells are the effectors of this arm of the immune system. Mature B cells produce antibodies against foreign antigens and can also act as antigen-presenting cells themselves (Rodriguez-Pinto 2005). Activation of the early B cell factor (*EBF*) gene is critical for the induction of B-cell development. *EBF* activates *PAX5* which acts as a positive regulator of *EBF*, enforcing a feedback loop. These two transcription factors then act in concert to inhibit T-lineage and myeloid gene expression programs.

While the model of lymphoid lineage specification is still evolving, the process of B-cell development from CLP to mature B-cell in humans is well established (Blom & Spits

2006). This progression, outlined in Figure 1.1, is marked by a number of immunophenotypic changes at each stage of development. These include differential expression of cluster of differentiation (CD) markers; CD34, CD10, CD127 (IL7R α), CD19, CD79a, and CD179a (Vpre-B) (Blom & Spits 2006). As cells progress through B-cell development, they can also be characterized by the presence or absence of specific immunoglobulin (*Ig*) gene rearrangements, which are important events in B lymphopoiesis. Expression of terminal deoxynucleotidyl transferase (*TdT*) and recombination activating gene (*RAG*) are both required for successful rearrangement of *Ig* loci and are expressed early in B-cell development (Corfe & Paige 2009). Prior to the immature B stage, development occurs exclusively within the bone marrow. Acquisition of surface IgM expression marks the stage at which immature B-cells exit the bone marrow and home to the spleen and other secondary lymphoid organs where they develop into mature B-cells.



CD34	+	+	+	-	-	-	-	-
CD10	+	+	+	+	+	+	+	-
IL-7R α	+	+	+	-	-	-	-	-
CD19	-	-	+	+	+	+	+	+
CD79a	-	+	+	+	+	+	+	+
TdT	-	-	+	-	-	-	-	-
RAG	-	-	+	+	-	+	+	-
Vpre-B	-	+	+	+	+	-	-	-
μ H	-	-	+/-	+	+	+	+	+
pre-BCR	-	-	-	-	+	-	-	-
IgH	GL	DJ _H	V _H DJ _H					
κ L	GL	GL	GL	GL	GL	V _L J _L	V _L J _L	V _L J _L
cycling	-	-	-	+	+	-	-	-
Pax-5	-	-	+	+	+	+	+	+
sIgM	-	-	-	-	-	-	+	+
sIgD	-	-	-	-	-	-	-	+

Figure 1.1 Model of human B-cell development

Adapted from Blom and Spits (2006).

1.2.2 Origins: Aetiology and incidence

Pre-B ALL has a peak incidence at 2-5 years of age (Gurney et al. 1995) and approximately one hundred and sixty new cases of ALL are diagnosed annually in Australia, with this number remaining unchanged over the last decade (Youlten et al. 2010). There is evidence that acquisition of genetic changes associated with ALL can occur *in utero* (Maia et al. 2003; Hjalgrim et al. 2002). This is supported by retrospective detection of genetic signatures associated with some forms of childhood ALL in Guthrie spot cards taken at birth from children who later developed ALL (Wiemels et al. 2009; Hjalgrim et al. 2002). Recurrent genetic abnormalities (discussed in detail later) have been associated with increased risk of developing ALL in early life and children with Down's syndrome (trisomy 21) are 10-15 times more likely to develop ALL, and they often suffer more treatment-related toxicities (Whitlock et al. 2005). Race and ethnicity has been shown to predict significantly different outcomes for children diagnosed with ALL, with children of African or Hispanic ancestry having a poorer prognosis than those with Anglo-Saxon ancestry (Bhatia 2004; McNeil et al. 2002), although this has been attributed by some to the lack of parity in healthcare for some minorities (Pui et al. 2003; Pui & Evans 2006). There is also a predilection towards a higher incidence of ALL in males, however the reason for this bias is unclear. *In utero* exposure to solvents, pesticides, hydrocarbons or ionizing radiation have all been linked to increased risk of developing ALL in early life (McNally & Parker 2006; Buffler et al. 2005), and studies on this topic are ongoing.

1.2.3 ALL therapy

Cancer in the very young can lead to significant problems for survivors in later life (Dowling et al. 2010). Cancer treatments in early life may result in impaired neurocognitive development and function (Ness et al. 2010), infertility, cardiovascular disease (Geenen et al. 2010), as well as impairments in musculoskeletal development (Ishida et al. 2010) and other complications. Furthermore, the experiences of cancer survivors can result in significant psychosocial disorders (Oeffinger et al. 2008). While current therapeutic regimens are largely successful, the present goal is to develop

therapies that are less toxic and will ultimately result in improved long term outcomes for survivors. Reducing the severity of treatment, while still attaining curative outcomes is important, thus balancing the type and amount of therapy with the risk of relapse is critical to reduce unnecessary exposure to toxic agents.

Despite the relatively high incidence of ALL in children, 5 yr event free survival (EFS) of up to 85-90% can now be achieved for some forms of the disease (Conter et al. 2010; Gaynon et al. 2010; Pui et al. 2010). Improved survival rates are attributable to the optimisation of complex chemotherapeutic regimens assessed in large scale, randomized clinical trials, as well as improved understanding of ALL biology (Pieters & Carroll 2010). Risk-based stratification of patients has also resulted in improved patient outcomes, and patients that are recognized as having a high risk of relapse are now receiving more intensive therapy. Unfortunately, those patients that do relapse often have a dismal prognosis, with relapsed ALL usually resistant to front-line chemotherapy.

Clinical management of ALL is continually undergoing improvements through large scale multicenter clinical trials run by major cooperative groups such as the Children's Oncology Group (COG) and the United Kingdom Children's Cancer Study Group (UKCCSG). These groups advise on best-practice treatment for a range of childhood cancers. Therapeutic protocols for the treatment of ALL endorsed by these organisations are similar in that they are comprised of several discreet stages, including induction therapy, consolidation, re-induction and maintenance therapy. However factors such as dosage, timing and the combination of drugs given to patients, differs among groups.

1.2.3.1 Induction therapy

Treatment for ALL incorporates several stages and in most cases continues for up to three years. The first stage of treatment is induction therapy and the primary goal of this phase is to restore normal haemopoiesis by eliminating as many leukaemic blasts as possible with high-dose chemotherapy. Three to four drugs are used depending on the perceived risk classification of patients (discussed later). Standard risk patients will

typically receive vincristine and asparaginase, as well as a glucocorticoid, either dexamethasone, prednisolone or prednisone (McNeer & Nachman 2010). Those patients that are identified as high risk may also receive an anthracycline, especially if they are over the age of 10 years (Pieters & Carroll 2010). The aim of induction chemotherapy is to achieve morphological remission in 4-6 weeks, and a patient's response to induction therapy has invariably been found to be the best predictor of long-term outcome.

The efficacy of induction therapy is evaluated by the detection of minimal residual disease (MRD) which can be defined as the level of leukaemic cells detectable in the bone marrow upon completion of induction therapy. MRD is used as a key indicator of response that guides the remainder of therapy during intensification and consolidation. The two main methods of MRD measurement are quantitative polymerase chain reaction (qPCR) and flow cytometry. The current limit of detection by flow cytometry is between 1 cell in 10^3 to 10^4 , while assessment of immunoglobulin (*IG*) or T-cell receptor (*TCR*) rearrangements by qPCR affords sensitivity of 1 cell in 10^4 to 10^5 (Bruggemann et al. 2010). While qPCR offers a superior level of sensitivity, this approach requires characterisation of rearrangements that are specific to the patients' leukaemic cells by sequencing the rearranged *IG* and *TCR* loci at diagnosis to develop specific primer/probe combinations. Leukaemic clones can continue to rearrange these loci during the progression of the disease, resulting in false-negatives which may inaccurately guide the remainder of therapy (Bruggemann et al. 2010; van der Velden & van Dongen 2009). Assessing a negative qPCR-MRD finding by flow cytometry offers confirmation, despite the obvious disparity in sensitivity between techniques.

1.2.3.2 Consolidation and re-induction

Once a negative MRD finding indicates that induction therapy has been effective in reducing the burden of leukaemic cells, patients begin the consolidation phase of treatment. Consolidation therapy aims to eliminate residual leukaemic cells that may be cycling slowly or have been driven into growth-arrest in response to chemotherapy. This phase typically includes asparaginase and high-dose methotrexate or 6-mercaptopurine.

Consolidation may also include intrathecal administration (injection into the cerebro-spinal fluid) of methotrexate with or without cytarabine and hydrocortisone to prevent central nervous system (CNS) relapse, particularly in high-risk patients (Pui & Evans 2006). Intrathecal administration of chemotherapy has been shown to have an equal or greater efficacy in preventing CNS relapse compared to cranial irradiation used by some groups (Stark et al. 2009), and introduction of intrathecal CNS prophylaxis has seen a substantial decrease in the practice of cranial radiation, which often has devastating long term effects on the developing brain. Upon completion of consolidation therapy, patients undergo re-induction therapy, sometimes referred to as delayed intensification therapy. This is essentially a repeat of the same therapy used in the induction phase and is a critical component of successful ALL treatment protocols (Pui & Evans 2006).

1.2.3.3 Maintenance therapy

At the completion of re-induction therapy, patients progress to maintenance therapy, which is the final stage of treatment. Maintenance therapy in paediatric ALL can continue for up to two years in females and three years in males. Standard maintenance therapy consists of daily 6-mercaptopurine and weekly methotrexate (Pieters & Carroll 2010). Some study groups have proposed the incorporation of pulsed administration of vincristine and dexamethasone, however the results of trials so far have not shown significant improvements in patient outcomes (Nagatoshi et al. 2010; Arico et al. 2008).

1.3 B-lineage acute lymphoblastic leukaemia

Although traditionally grouped together, cases of B-lineage ALL can be further delineated into different subtypes based on the immunophenotype of leukaemic cells. These subtypes are associated with differentiation stages of B-cell development and have significant prognostic ramifications (Pieters & Carroll 2010; Bene 2005). At the time of diagnosis, patient samples are assessed for immunophenotype and are grouped into one of three types; early pro-B, pre-B and transitional pre-B, while mature B-cell ALL is rarely diagnosed in children (Pui 2006). Table 1.1 outlines these subtypes, and

their immunological features and frequency. It is important to note that there is a degree of plasticity in the immunophenotype of B-lineage ALL cells, making a determination as to the specific subtype difficult in some cases. For this reason, B-lineage ALL is frequently referred to as precursor B-cell ALL (pre-B ALL) as a whole, and often more importance is placed on identifying specific genomic alterations associated with clinical outcomes, as these are considered more important for risk-based stratification.

Table 1.1 Immunologic classification of B-lineage ALL

Table adapted from Childhood Leukemias by Ching-Hon Pui (Pui 2006).

Subtype	Immunological marker (% pos)								frequency
	CD19	CD20	CD22	CD79a	CD10	cIgμ	sIgμ	sIg κ or λ	
Early pre-B (pro-B)	100	35	99	99	95	0	0	0	60-65%
Pre-B	100	45	100	100	100	100	0	0	20-25%
Transitional pre-B (late pre-B)	100	55	100	100	100	100	100	0	10-12%
Mature B	100	99	100	99	50	100	~95	~95	3-5%

Abbreviations: cIgμ, cytoplasmic immunoglobulin mu heavy chain; sIgμ, surface immunoglobulin mu heavy chain; sIgκ or λ, surface immunoglobulin kappa or lambda light chain; CD, cluster of differentiation.

1.3.1 Prognosis and risk stratification

As stated previously, vast improvements have been achieved in the long term EFS of paediatric ALL patients. A cornerstone of contemporary ALL management is the risk-based stratification of patients based on presenting features (Pui et al. 2011). By directing patients expected to have a poor outcome to more intensive therapy, these high-risk patients are afforded the best possible chance of disease clearance. Furthermore, by limiting the amount of therapy given to patients identified as having a favourable outcome, these low-risk patients are spared the detrimental short and long-term side effects associated with agents used in high-risk therapy (Pui & Evans 2006).

Clinical features that predict outcome include age, gender and white blood cell (WBC) count at diagnosis. Biological factors with prognostic relevance include immunophenotype (pro-B, pre-B or common B) and the genotype of the patient's

leukaemic cells. Hyperdiploidy defined as > 51 chromosomes or a DNA index greater than 1.16 is associated with a favourable outcome in standard risk pre-B ALL (Heerema et al. 2000; Sutcliffe et al. 2005). Several recurrent cytogenetic abnormalities are associated with ALL biology and are discussed in the next section. Table 1.2 shows the clinical and biological features that are used to initially stratify patients diagnosed with pre-B ALL as either standard or high-risk. Features associated with a higher risk include age of 0-1yr or greater than 9yrs, male gender, high WBC count ($> 50 \times 10^9/L$), pro-B immunophenotype and hypodiploidy defined as less than 45 chromosomes. As discussed earlier, perhaps the most important prognostic factor in ALL is the response to induction therapy detected as MRD. This is a strong predictor of relapse or resistance to therapy and determines the appropriate treatment protocol for patients, either intensification or consolidation of treatment administered.

Table 1.2 Clinical and biological features for initial patient stratification

Adapted from Pieters and Carroll (2010).

Factor	Standard Risk	High Risk
Age at diagnosis	1-9 yrs	< 1yr or > 9 yrs
Sex	Female	Male
WBC count	$< 50 \times 10^9 / L$	$> 50 \times 10^9 / L$
Immunophenotype	precursor B ALL	B ALL
Genotype	Hyperdiploid >50 chromosomes	hypodiploidy < 45 chromosomes
Post-induction MRD	no blasts detected	detectable blasts

1.3.2 Molecular types and cytogenetics

Numerous recurrent cytogenetic anomalies have been characterised in pre-B ALL and these are associated with different outcomes (Pui & Evans 2006). Fusion transcripts generated by some chromosomal rearrangements have been demonstrated to have leukaemogenic potential. Juxtaposition of genes encoding developmentally critical DNA-binding proteins with foreign loci can result in the production of fusion proteins

with unrestricted and, or, oncogenic transcriptional activity (Poppe et al. 2007). A high proportion of cytogenetic rearrangements associated with leukaemia involve transcription factors required for, or involved in, lymphoid development. Some examples of these are the master transcriptional regulator mixed lineage leukaemia (*MLL*) or the runt-related transcription factor 1 (*RUNX1* also known as *AML1*). The most common cytogenetic alterations and their prognostic implications are outlined in Table 1.3.

1.3.2.1 *TEL-AML1*

The *TEL-AML1* fusion (also known as *ETV6-RUNX1*) is characterised by the t(12;21)(p13;q22) translocation and is the most common cytogenetic rearrangement in paediatric pre-B ALL, occurring in around 25% of cases (Onciu 2009). *TEL-AML1* is associated with high chemosensitivity, and thus a favourable outcome. Both *TEL/ETV6* and *AML1/RUNX1* are transcription factors with established roles in haemopoiesis (Okumura et al. 2007; Wang et al. 1998). Targeted disruption of *TEL* after the establishment of haemopoiesis results in an accumulation of progenitor cells at the pro-B stage, however the developmental potential of these cells is not blocked completely (Fischer et al. 2005). These data reflect what is seen in human disease, as carriers of the *TEL-AML1* rearrangement do not always progress to overt leukaemia and identical twin studies have demonstrated that it is possible to carry *TEL-AML1* in haemopoietic cells without any manifestation of disease (Maia et al. 2003). In studies conducted by Greaves and colleagues, *TEL-AML1*-transduced human CD34^{pos} cord blood cells, transplanted into NOD-SCID mice, gave rise to cells that display B-cell developmental potential and also exhibit greatly enhanced resistance to some apoptotic stimuli, including Fas ligand, camptothecin and melphalan, suggesting that these pre-leukaemic clones have a growth advantage over non-transduced cells (Hong et al. 2008). In carriers of *TEL-AML1*, progenitor cells harbouring this feature are referred to as pre-leukaemic clones and tend to display an early pre-B/pro-B immunophenotype. It is postulated that they require a second hit to become overtly tumorigenic (Greaves 2009).

Table 1.3 Common cytogenetic abnormalities in pre-B ALL

Adapted from Onciu, M. (2009).

Cytogenetic group	Frequency (%)	Fusion gene	Cytogenetic abnormality	Prognosis	Notes
Hyperdiploid	27-29	NA	51-65 chromosomes	Low-risk	Higher sensitivity to MTX, MP
ALL with t(12;21)	25	<i>TEL-AML1</i> (<i>ETVX-RUNX1</i>)	t(12;21)(p13;q22)	Low-risk	Higher sensitivity to L-asparaginase
Hypodiploid	5-6	NA	<45 chromosomes	High-risk	B-cell developmental gene mutation in 100% of cases
ALL with t(1;19)	3-6	<i>TCF3-PBX1</i> (<i>E2A-PBX1</i>)	t(1;19)(q23;p13)	Standard-risk	NA
Philadelphia positive ALL	3-5	<i>BCR-ABL</i>	t(9;22)(q34;q11.2)	High-risk	<i>IKZF1</i> (Ikaros) deletions common
ALL with <i>MLL</i> (11q23) rearranged	2-3	<i>AF4-MLL</i> <i>ELL-MLL</i>	t(4;11)(q21;q23) t(19;11)(p13;q23)	High-risk High-risk	Increased expression of <i>HOX</i> genes.

1.3.2.2 *TCF3-PBX1*

The t(1;19)(q23;p13) *TCF3-PBX1* (also known as *E2A-PBX1*) translocation occurs in around 5% of cases of paediatric B-lineage ALL. Expression of this fusion gene produces an oncogenic protein containing the N-terminal transactivation domain encoded by the transcription factor 3 gene (*TCF3*) fused to the DNA-binding homeodomain encoded by pre-B-cell leukaemia homeobox 1 gene (*PBX1*). The resulting fusion protein can induce expression of genes defined by the *PBX1* DNA-binding domain alone (Van Dijk et al. 1993) or alternatively the *TCF3-PBX1* oncoprotein can dimerize with *HOX* proteins, resulting in enhanced expression of *HOX* target genes (Aspland et al. 2001). *TCF3* encodes two basic helix-loop-helix (bHLH) proteins, E12 and E47 that form homodimeric complexes to regulate immunoglobulin gene expression (Bain et al. 1999). The *TCF3* gene is required for lymphoid development and disruption of normal *TCF3* expression results in accumulation of cells at the pro-B cell stage (Uckun et al. 1998b). Recent studies investigating conditional knockout of *Tcf3* in murine embryonic stem cells have demonstrated that *Tcf3* gene expression plays a critical role in repressing expression of genes associated with self-renewal such as *Nanog*, *Tcl1*, *Tbx3* and *Esrrb* (Yi et al. 2008). Dysregulated expression of the *TCF3-PBX1* may result in improper activation of gene expression programs associated with self-renewal.

The t(1;19) feature was historically associated with a poor outcome, however improvements in multi-agent therapy have resulted in this rearrangement being removed from risk stratification matrices (Raimondi et al. 1990). This may need to be reviewed however, as a recent study has identified an association between *TCF3-PBX1* and CNS relapse in children with pre-B ALL (Jeha et al. 2009). Re-inclusion of this genotype in clinical matrices would direct these patients to intensified CNS-directed prophylaxis which may have a significant effect on patient outcomes.

1.3.2.3 *BCR-ABL1*

Perhaps one of the best-known translocations associated with leukaemia is the t(9;22)(q34;q11.2) *BCR-ABL1* fusion, also known as the Philadelphia chromosome. This rearrangement is characterised by translocation of the coding region of *ABL1* from chromosome 9 to the *BCR* gene on chromosome 22 and occurs in 3-5% of cases of childhood ALL (Arico et al. 2000). This event results in constitutive expression of the *BCR-ABL1* fusion RNA driven by the *BCR* promoter (Pieters & Carroll 2010). While the precise function of the *BCR* gene has not been established, *ABL1* is a proto-oncogene encoding a non-receptor tyrosine kinase and the *BCR-ABL1* fusion protein has been demonstrated as a driver of oncogenesis (Uckun et al. 1998a). Transformation by *BCR-ABL1* activates a number of intracellular signalling pathways, and those shown to play major roles in *BCR-ABL1* induced cellular proliferation and transformation include activation of RAS, PI-3 kinase and Jun kinase (Pui 2006). *BCR-ABL1* transformation also enhances resistance to some cytostatic drugs and radiation by upregulating the anti-apoptotic *BCL2* family member BCL-XL (Skorski 2002). Children with ALL positive for the *BCR-ABL1* genotype usually have a high MRD after induction therapy and this translocation is associated with high-risk and a poor outcome with 5ys EFS of around 28% (Arico et al. 2000).

BCR-ABL1-positive ALL frequently harbours deletions in B-cell developmental genes (Onciu 2009). Recent data gathered on *BCR-ABL1*-positive ALL indicated that 75% of children and 91% of adults harboured mutation of at least one *IKAROS* allele (Mullighan et al. 2008). Iacobucci and colleagues recently reported the acquisition of mutations in *IKAROS* splice sites after treatment with tyrosine kinase inhibitors, suggesting a role for dominant-negative *IKAROS* isoforms in the acquisition of Imatanib and Dasatanib resistance in *BCR-ABL1*-positive patients (Iacobucci et al. 2008).

1.3.2.4 *MLL* rearrangements

Translocations involving the mixed lineage leukaemia (*MLL*) gene at 11q23 are very common in infant ALL (< 1 yr) with around 80% of such cases attributable to

rearrangements of the *MLL* locus. In contrast, only 2% of children aged 1-9 yr present with *MLL* rearrangements (Somerville & Cleary 2010). Universally, *MLL* involvement is associated with high-risk and a poor outcome. Leukaemic cells containing *MLL* rearrangements are usually highly resistant to glucocorticoids and L-asparaginase, two front-line chemotherapeutic agents used in standard induction therapy (Pieters & Carroll 2010). *MLL*-rearranged ALL exhibits an early pro-B immunophenotype and CD10 negativity and mutations in B-cell developmental genes are common (Ayton & Cleary 2001). The striking feature of *MLL* rearrangements is the large number of partner loci involved in 11q23 translocations. To date over 100 different *MLL* rearrangements have been described (Meyer et al. 2009).

MLL is required for normal haemopoiesis (Jude et al. 2007), however the precise mechanism by which *MLL* rearrangements induce transformation is unclear. *MLL* encodes a large, multidomain protein (3972 amino acids), that contains an AT hook DNA-binding domain, as well as a H3K4 histone methyltransferase moiety which is associated with epigenetic activation of homeobox genes by chromatin remodelling. Regardless of the 11q23 fusion partner, *MLL*-fusion oncoproteins consistently induce high level transcriptional activation of *HOXA* and *MEIS1* genes through epigenetic activation of these gene loci (Somerville & Cleary 2010). These genes are normally repressed during terminal differentiation and constitutive expression is thought to contribute to uncontrolled proliferation of leukaemic cells harbouring *MLL* fusions.

1.3.2.5 Numerical chromosomal alterations

Changes in chromosome number (ploidy) are frequently observed in paediatric ALL. Hyperdiploidy is defined as greater than 50 chromosomes and is the most common genetic abnormality present in childhood ALL with an incidence of around 25-30% (Onciu 2009). Hyperdiploidy confers a favourable prognosis and current estimates of overall survival of children with hyperdiploid ALL is 90% (Paulsson & Johansson 2009). Furthermore, cases with trisomies 4, 10 and 17 have a particularly favourable outcome (Sutcliffe et al. 2005). Activating mutations affecting *FLT3* occur in around

20% of hyperdiploid ALL and efforts are underway to incorporate small molecule inhibitors of *FLT3* in standard therapy as these have been shown to be highly effective at killing *FLT3* positive, hyperdiploid ALL (Brown et al. 2005).

Hypodiploidy which is defined as fewer than 45 chromosomes, is much less frequent in childhood ALL, occurring in around 5% of cases of paediatric B-lineage ALL (Onciu 2009). Due to its low frequency, there have been few studies investigating the biological and clinical significance of this genetic subtype of ALL, however patients harbouring this genetic abnormality have a dismal outcome. A study conducted by Harrison and colleagues found that the 3yr EFS of 121 children with 42-45 chromosomes was around 65%, while those with 25-39 chromosomes (n=20) exhibited a 3yr EFS of only 30% (Harrison et al. 2004). As one might expect, deletions of B-cell developmental genes occurred in 100% percent of cases (Harrison et al. 2004).

1.3.2.6 Recurrent mutations in pre-B ALL

Genome-wide screens have provided insights into the molecular pathogenesis of ALL (Mullighan et al. 2007; Mullighan & Downing 2009). The most commonly disrupted loci in pre-B ALL are those encoding transcription factors involved in the regulation of B-cell development such as *PAX5*, *IKZF1* (*IKAROS*), *IKZF2* (*HELIOS*), *IKZF3* (*AIOLOS*), *LEF1*, *TCF3* and *BLNK*. Perturbations in lymphoid developmental genes occur in around 60% of cases of pre-B ALL and explain the developmental blockade that is a feature of this disease (Kuiper et al. 2007b; Mullighan et al. 2007). Surprisingly, DNA damage response pathways and tumour suppressor genes that are often mutated or lost in other tumour types are normally intact in pre-B ALL, with the prominent exception of the p16^{*INK4*}/*p14*^{*ARF*} locus which is altered in around 30% of paediatric pre-B ALLs (Kim et al. 2009).

Loss of function mutations affecting *IKAROS* (*IKZF1*) are associated with haematological malignancies including AML (Yagi et al. 2002) and pre-B ALL (Mullighan et al. 2007), and inactivating mutations or deletions affecting the *IKAROS*

locus are highly predictive of relapse in paediatric pre-B ALL (Kuiper et al. 2010; Mullighan et al. 2009). Ikaros proteins are nuclear factors required for normal haemopoiesis and lymphoid specification (Koipally et al. 1999; Cobb & Smale 2005). Mice homozygous for germline mutations in the murine Ikaros DNA-binding domain completely lack lymphoid and natural killer cells (Georgopoulos et al. 1994), while retaining normal development within the myeloid and erythroid compartment.

Ikaros proteins are sequence-specific transcription factors that have also been implicated in chromatin remodelling and can both repress and potentiate gene expression (Koipally et al. 1999; Rebollo & Schmitt 2003). Ikaros family proteins can form homodimers or heterodimers with other Ikaros family members (Rebollo & Schmitt 2003; Georgopoulos et al. 1994) and isoforms of the IKAROS protein generated by alternative splicing have been identified in both normal haemopoietic cells and leukaemic blasts (Tabayashi et al. 2007; Rebollo & Schmitt 2003; Iacobucci et al. 2008; Meleshko et al. 2008). These isoforms frequently lack part or all of the N-terminal zinc finger DNA-binding domain. However, the C-terminal dimerisation domain typically remains unaffected, thus these alternative isoforms can bind to canonical IKAROS (or other Ikaros family proteins) and act as dominant negative (DN) inhibitors (Sun et al. 1996).

1.3.2.7 Classical oncogenes associated with pre-B ALL

As well as activation of proto-oncogenes through cytogenetic rearrangements that are unique to haematological malignancies, there are also a number of “classical” oncogenes involved in leukaemogenesis, particularly in the case of *TEL-AML1* carriers who are thought to require a second oncogenic hit to progress to overt leukaemogenesis (Hong et al. 2008). One such example is the B-cell CLL/lymphoma gene (*BCL2*) (Vaux et al. 1988). *BCL2* activation is most frequently associated with t(14;18) translocations, however *BCL2* can also be overexpressed by activating point mutations or gene amplification. *BCL2* protein inhibits apoptosis by binding to and inhibiting the pro-apoptotic protein BCL2-associated X protein (BAX) (Zivny et al. 2010). While

expression of *BCL2* in ALL is common, it has not been associated with outcome (Gala et al. 1994).

One of the best studied oncogenes is the v-myc myelocytomatosis viral oncogene homolog, better known as *MYC*, which is overexpressed in a large number of solid tumours as well as some leukaemias (Albihn et al. 2010; Rubnitz & Crist 1997). T(8;14) rearrangements result in the translocation of the *MYC* gene to the immunoglobulin heavy chain (*IgH*) locus, resulting in an overexpression and an abundance of MYC protein driven by the *IgH* promoter.

MYC activity requires dimerisation with the basic helix-loop-helix (bHLH) protein MYC-associated factor X (MAX). The MYC-MAX dimer modulates gene transcription by binding to hexanucleotide sequences termed E boxes 5'CA[C/T]GTG-3' in the promoter of target genes (Luscher & Larsson 1999). MYC also interacts with proteins involved in chromatin remodelling such as transformation/transcription domain-associated protein (TRRAP), RuvB-like 1 (RUVBL1/TIP49) and RuvB-like 2 (RUVBL2/TIP48) (Amati et al. 2001; Wood et al. 2000; Park et al. 2001). A large number of genes are targets of MYC and many of these have been documented in the MYC Target Gene Database available online at: <http://www.myc-cancer-gene.org/site/mycTargetDB.asp>. To date this database contains 1697 individual genes that are predicted to be targets of MYC, and studies are ongoing to establish patterns of gene expression associated with oncogenic MYC activation (Albihn et al. 2010).

1.3.2.8 Tumour suppressor genes in pre-B ALL

Disruption of tumour suppressor genes (TSGs) are infrequent events in childhood ALL with the exception of those affecting the p16^{INK4}/p14^{ARF} locus which is altered in 25-30% of paediatric B-lineage ALLs by either mutation or hypermethylation (Krug et al. 2002; Kim et al. 2009). This locus encodes the cyclin-dependent kinase inhibitor 2A (*CDKN2A*) gene which can give rise to several different mRNA transcripts. Two *CDKN2A* variants encode protein isoforms of p16 that inhibit CDK4 kinase, while a

third *CDKN2A* transcript contains an alternative open reading frame (ARF) encoding the p14 protein which can inhibit MDM2. MDM2 binds to and induces degradation of the tumour suppressor p53, and p14 inhibits this interaction (Pomerantz et al. 1998). CDK4 and p53 both have prominent roles in regulating transition from G1 to S phase and are tightly regulated. Mutations involving the gene encoding p53 (*TP53*) are very common in solid malignancies, however the incidence of p53 mutations is extremely low in childhood cancers in general. Disruption of p53 is observed in around 5% of paediatric ALL cases, typically correlating with a poor outcome (Marks et al. 1997).

1.4 *CTGF* gene expression is dysregulated in pre-B ALL

A microarray study performed by Boag and colleagues compared the global gene expression profiles of 102 paediatric ALL samples with two control populations, bone marrow derived CD34^{pos} cells representing the HSC to pro-B cell compartment, and CD19^{pos}/IgM^{neg} cells representing the pro-B to small pre B-II compartment. It was expected that this approach would identify genes that may be relevant to leukaemogenesis (Boag et al. 2007). The results of this study demonstrated that one gene in particular, connective tissue growth factor (*CTGF*) displayed elevated expression, 19-fold higher vs CD34^{pos} and 37-fold higher vs CD19^{pos} IgM^{neg} cells in 75% of pre-B ALL samples. Expression was restricted to pre-B ALL and *CTGF* expression was not detected in T-ALL cells. Curiously, *CTGF* expression was also absent in the *TCF3-PBX1* cytogenetic group of pre-B ALL. Importantly, these findings were validated in an independent cohort of 132 paediatric ALL patients used in a study at St Jude Children's Research Hospital, Memphis (Ross et al. 2003). Deregulation of *CTGF* in B-lineage ALL has now been identified in four independent studies (Vorwerk et al. 2000; Salazar-Torra et al. 2007; Boag et al. 2007; Kang et al. 2010). *CTGF* is not normally expressed by haemopoietic cells at any stage of development and has no known role in lymphocyte biology, thus the role of *CTGF* in pre-B ALL and the mechanism promoting its aberrant expression are unclear and remain to be elucidated.

1.5 Connective tissue growth factor

The name connective tissue growth factor (CTGF) was coined in 1991 by Bradham and colleagues to describe a novel protein secreted by a human umbilical endothelial cell line (HUVEC) that stimulated mitogenesis and chemotaxis of murine fibroblasts *in vitro* (Bradham et al. 1991). Subsequently, overexpression of CTGF has been observed in a number of fibrotic pathologies (Manetti et al. 2007; Lang et al. 2007; Leask 2004; Faull 2005). CTGF also has roles in numerous other biological and developmental processes including wound healing, angiogenesis, chondrogenesis, adhesion, apoptosis, chemotaxis, mitogenesis, extracellular matrix formation, endochondral ossification and tumorigenesis (Leask & Abraham 2006; Brigstock 2003; Dhar & Ray). CTGF appears to exert its effects through interacting with other protein factors through one of its four functional domains or by interacting with cell surface receptors and integrins. While some of these interactions have been characterised in various cell types, the precise role of CTGF in these biological contexts is largely unknown. Several recent reviews on CTGF function have highlighted the role of CTGF as a matricellular protein rather than a classical growth factor as CTGF is thought to facilitate interactions between cells and the surrounding extracellular matrix (Chen & Lau 2009; Brigstock 2010; Leask 2010). There is a growing body of evidence that CTGF acts to connect cells to the microenvironment by providing adhesive signals as well as modulating the availability of soluble growth factors, thus the tissue context within which CTGF is expressed is likely to play a key role in determining its biological function in any given microenvironment.

1.5.1 Origins

CTGF also known as CCN2, is a member of the CCN family of structurally related proteins, named after the three founding members Cysteine rich 61 (Cyr61 or CCN1), connective tissue growth factor (CTGF) and nephroblastoma overexpressed (NOV or CCN3) (Brigstock 2003). The remaining CCN members include the WNT1 inducible signalling pathway protein (WISP) members 1 to 3; WISP1 (CCN4), WISP2 (CCN5) and WISP3 (CCN6) (Leask & Abraham 2006). All six proteins share a similar modular

structure, as illustrated in Figure 1.2, containing up to four highly conserved domains. These domains contain motifs that resemble functional elements present in important regulatory proteins, including the insulin-like growth factor binding proteins (IGFBP), von Willebrand type-C (VWC), thrombospondin type-1 (TSP1) and a cysteine knot motif in the CT domain (Perbal 2001; Leask & Abraham 2006). This curious mosaic structure has resulted in CTGF (and other CCN proteins) being implicated both directly and indirectly in a vast array of biological processes and has made predicting the function of CTGF in any given biological context difficult (Chen & Lau 2009; Holbourn et al. 2008).

1.5.2 Protein structure

The CTGF protein contains 349 amino acid residues and has a molecular weight of 36-38 kDa depending on glycosylation of the full length protein (Bradham et al. 1991; Ball et al. 2003b). CTGF is a secreted protein and the N-terminus contains a 37 amino acid signal peptide (SP) directing the protein through the Golgi apparatus to the cell membrane for secretion (Abreu et al. 2002; Chen et al. 2001). CTGF is thought to have a short half-life and is susceptible to proteolysis by several enzymes, including elastase, plasmin and the matrix metalloproteases (MMPs) 1, 3, 7 and 13 (Hashimoto et al. 2002). Several smaller molecular weight fragments of CTGF have been detected in biological fluids and these are thought to potentiate specific biological functions (Yang et al. 1998; Steffen et al. 1998; Ball et al. 1998). These smaller isoforms, which arise from proteolysis upon secretion, range in size from 10-20 kDa and are primarily comprised of either module 4 alone or modules 3 and 4 combined (Brigstock et al. 1997). The two enzymes elastase and plasmin can cleave CTGF between any of the four modules (Hashimoto et al. 2002), while MMPs 1, 3, 7 and 13 cleave the CTGF protein in the proteinase-sensitive hinge region between modules 2 and 3 (Gressner & Gressner 2008).

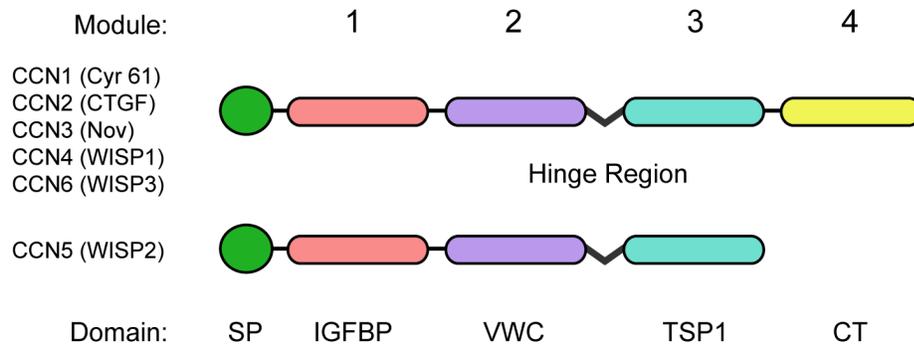


Figure 1.2 Modular structure of CCN proteins

All CCN members share a similar modular structure including an N-terminal signal peptide (SP), insulin-like growth factor binding protein domain (IGFBP), Von Willebrand type-C (VWC) domain, and a thrombospondin type-1 domain (TSP1). CCNs1-4 and CCN6 contain a cysteine-rich C-terminus (CT) which is not present in CCN5. All CCN family members contain a protease-sensitive hinge region between module 2 and 3. Figure adapted from Leask, A. and Abraham, D. (2006).

1.5.3 CTGF signalling

CTGF is pleiotropic and has been implicated in numerous biological processes, largely due to the diversity of proteins shown to interact with CTGF and other CCN family members. Our understanding of the various roles CTGF may play in any given microenvironment or biological process is limited at present and important questions that require addressing include which cellular receptors can be activated by CTGF, which soluble factors can CTGF bind to, and can these interactions modulate the function of target proteins? The following sections outline some of the key protein factors and receptors that are known to bind CTGF.

1.5.3.1 Insulin signalling

Module 1 of CTGF contains a sequence motif homologous to the IGFBP superfamily and this domain has been shown to bind human recombinant insulin-like growth factor (IGF) 1 and 2 with low affinity (Kim et al. 1997). IGF family proteins have been

implicated in haemopoietic development and importantly in leukaemia (Zumkeller 2002; Elmlinger et al. 1996). Furthermore, the first observation that *CTGF* was expressed specifically in pre-B ALL came from a study investigating the expression of insulin binding protein genes in leukaemia (Vorwerk et al. 2000). However, there is little functional evidence so far that CTGF plays a significant role in IGF signal transduction and the biological implications of the structural relationship between CTGF and IGFbps remains unclear and warrants further investigation.

1.5.3.2 TGFB family signalling

CTGF module 2 contains a cysteine rich VWC motif and this domain can bind members of the transforming growth factor beta superfamily of proteins including; bone morphogenic proteins (BMPs) 4 and 7 and transforming growth factor beta 1 (TGFB1) (Abreu et al. 2002; Nguyen et al. 2008). Binding modulates the availability of these factors to attach to their cognate receptors and activate intracellular signalling. Interaction between CTGF and either BMP4 or BMP7 impairs the ability of these proteins to bind to the type II TGF receptor (TGFBR2). Conversely, TGFB1 bound to CTGF displays enhanced TGFBR2 binding (Abreu et al. 2002). Interestingly, CTGF can be potently induced by TGFB1 signalling and this is the best studied inducer of *CTGF* gene expression (Trojanowska 2009; Arnott et al. 2008). Furthermore, around a third of TGFB1-responsive genes are no longer inducible in *CTGF* null embryonic fibroblasts, demonstrating the critical role CTGF plays in signalling downstream of TGFB1 (Shi-wen et al. 2006). These data indicate that the CTGF/TGFB1 axis may demonstrate a positive feedback loop.

1.5.3.4 Vascular endothelial growth factor signalling

CTGF has been suggested to play a role in regulating the angiogenic effects of vascular endothelial growth factor (VEGF). Inoki and colleagues identified CTGF in a screen for VEGF-binding partners using a yeast two-hybrid system (Inoki et al. 2002). The authors further demonstrated that CTGF has the ability to bind to VEGF, using either module 3

(TSP1 domain) or module 4 (CT domain) independently. Full length CTGF sequesters VEGF in a complex preventing it from binding to cell surface receptors, and experiments have shown that cleavage of this complex by a number of MMPs in the CTGF hinge region can result in the release of previously bound VEGF (Hashimoto et al. 2002; Ito et al. 2007). This mechanism is suggested to fine-tune the bioavailability of VEGF. Furthermore, MMPs are commonly expressed and activated during wound healing and tissue remodelling (Murphy & Nagase 2008), two environments where angiogenesis is required. An overall increase in cell surface or extracellular expression of MMP family proteins is likely to impair the ability of CTGF to bind VEGF and induce the release of any VEGF that is sequestered in either stromal or cell surface-bound complexes. Curiously, reminiscent of TGF β 1, induction of *CTGF* gene expression by VEGF has also been reported (He et al. 2003; Suzuma et al. 2000; Kuiper et al. 2007a), however these findings were obtained in studies of bovine retinopathies and may not represent the role for CTGF in human biology, and further investigation is required.

1.5.4 CTGF cell surface interactions

As well as binding important regulatory proteins, CTGF has been shown to bind several membrane-bound receptors that can facilitate intracellular signalling *via* tyrosine phosphorylation. Because the response to extracellular CTGF appears to vary depending on the tissue context, an important step in elucidating the potential role of CTGF in any given microenvironment is to determine what receptors are available to bind CTGF and what cells are expressing such receptors. Some of the known CTGF receptor interactions that have been validated are outlined below.

1.5.4.1 LRP1 receptor

A study performed by Segarini and colleagues demonstrated the ability of CTGF to bind the low density lipoprotein receptor-related protein (LRP1), also known as α 2-macroglobulin receptor (α ₂MR) through interactions with the TSP1 domain in module 3

of the CTGF protein (Segarini et al. 2001). In this study, the authors used LRP1 purified from lysates prepared from the bone marrow stromal cell line BSM2, suggesting that LRP1 has the potential to act as a CTGF receptor in the bone marrow microenvironment. LRP1-CTGF interactions have also been confirmed in rat hepatic stellate cells and human renal fibroblasts (Gao & Brigstock 2003; Yang et al. 2004). Yang and colleagues confirmed phosphorylation of the intracellular domain of LRP1 in response to CTGF binding, resulting in activation of downstream mitogen-activated protein kinase (MAPK) signalling (Yang et al. 2004). Blockade of LRP1 with receptor associated protein (RAP) or inhibition of MAPK activation with the MAPK/ERK kinase 1 (MEK1) inhibitor PD98059 reduced myofibroblast differentiation in response to exogenous CTGF (Yang et al. 2004). CTGF has also been shown to bind the Wnt-co receptor LRP6 on *Xenopus* embryos, inhibiting Wnt signaling. However, this has not been investigated in mammalian cells (Mercurio et al. 2004).

1.5.4.2 NTRK1/NGFR receptor complex

CTGF is thought to play a role in diabetic retinopathy and a study by Wahab et al demonstrated that CTGF activates signalling pathways, including ERK 1/2, JNK, ERK5, phosphatidylinositol 3-K, CaM-KII, PKC α , and PKC δ in human mesangial cells (Wahab et al. 2005). Furthermore they demonstrated that this response was downstream from CTGF binding to the type 1 neurotrophic tyrosine kinase receptor (NTRK1/TrKA) and the co-receptor nerve growth factor receptor (NGFR/p75^{NTR}). The cytoplasmic domain of NTRK1 contains tyrosine residues that can be phosphorylated and interact directly with intracellular adapter proteins, resulting in a vast array of biological responses, including proliferation, cytoskeletal remodelling and membrane trafficking, and differentiation (Huang & Reichardt 2003).

1.5.4.3 CTGF-associated adhesion

Module 4 of CTGF (CT domain) binds heparin sulphate proteoglycans (HSPGs). These proteins are expressed on the surface of some cell types and HSPG-CTGF interactions

have been demonstrated in fibroblasts (Chen et al. 2000), hepatic stellate cells (Gao & Brigstock 2004), and some monocytes (Schober et al. 2002). In each case, specific integrin complexes were required as co-receptors and these differed between cell types, including $\alpha\text{m}\beta\text{2}$ integrins in monocytes, $\alpha\text{6}\beta\text{1}$ in fibroblasts and $\alpha\text{5}\beta\text{3}$ in hepatic stellate cells. Furthermore binding of CTGF to $\alpha\text{5}\beta\text{3}$ integrins through the CT domain facilitates migration and adhesion of chondrosarcoma cells independent of HSPG binding (Tan et al. 2009). Other integrins known to interact with CTGF include $\alpha\text{5}\beta\text{1}$ on pancreatic stellate cells and chondrocytes (Gao & Brigstock 2006; Hoshijima et al. 2006; Nishida et al. 2007), and $\alpha\text{IIb}\beta\text{3}$ on human platelets (Jedsadayamata et al. 1999). Hoshijima and colleagues have demonstrated that fibronectin (FN1) also binds the CT domain, promoting adhesion of chondrocytes. This process was dependant on co-binding $\alpha\text{5}\beta\text{1}$ integrins (Hoshijima et al. 2006).

HSPGs are common components of stroma and this has led to the hypothesis that the microenvironment may act as a sink for bioactive CTGF. Perlecan is the major extracellular matrix proteoglycan in the bone marrow (Schofield et al. 1999) and has been shown to sequester CTGF as well as a number of other important protein factors, including fibroblast growth factors (FGFs), platelet derived growth factor (PDGF), VEGF and many more (Melrose et al. 2008). It is not known whether CTGF bound to stroma can interact with other binding partners such as TGFB1, BMPs or VEGF and careful functional characterisation of HSPG-bound CTGF is required before such fundamental biological questions can be answered.

1.5.5 The *CTGF* locus

The human *CTGF* gene maps to the long arm of chromosome 6, at cytoband q23.1 (Martinerie et al. 1992). *CTGF* is an immediate early gene and can be expressed rapidly in the absence of endogenous protein synthesis, highlighting its importance in mammalian homeostasis (Igarashi et al. 1993). The *CTGF* gene is relatively small at approximately 3.5 kb in length and encodes a single mature mRNA transcript of 2.4 kb in length, containing five exons. There have been no conclusive reports of *CTGF* splicing variants or alternative transcripts, however a recent study investigating the

exome of thymic tumours identified *CTGF* mRNA transcripts with reduced inclusion of exon 2 (Soreq et al. 2008). Follow-up studies need to be performed to accurately characterize *CTGF* mRNA transcripts in these samples.

1.5.5.1 The *CTGF* promoter

CTGF expression can be modulated by a number of extracellular stimuli, including high glucose (Wang et al. 2008), hypoxia (Higgins et al. 2004; Hong et al. 2006), mechanical stress (Nishida et al. 2008) and soluble factors such as transforming growth factor beta (TGFB1) (Xie et al. 2005; Arnott et al. 2008), endothelin-1 (ET-1) (Recchia et al. 2009), serum response factor (SRF) (Muehlich et al. 2007) and curiously the matrix metalloprotease MMP3 (Eguchi et al. 2008). Other factors have also been implicated in the induction of *CTGF* expression, however these observations are largely cell type specific and in some cases indirect, requiring further investigation. The *CTGF* promoter contains several well characterised regulatory elements, outlined in Figure 1.3. These include binding sites for Sp1 and Ap-1/c-Jun (Holmes et al. 2003; Yu et al. 2009), tandem Ets-1 binding motifs (Van Beek et al. 2006), as well as a Smad-binding element (SBE) (Holmes et al. 2001), and the basal control element (BCE-1), originally named the TGFB1 response element (TBRE) (Grotendorst et al. 1996).

TGFB1 has long been recognised as one of the most potent inducers of *CTGF* expression in fibroblasts, mesangial cells and vascular smooth muscle cells (Grotendorst 1997; Chen et al. 2002; Fu et al. 2001). The BCE-1 was originally described by Grotendorst and colleagues as a promoter region critical for the induction of *CTGF* expression by TGFB1. This promoter element is activated downstream of MEK/ERK signalling and mutation of this region can abolish TGFB1-induced *CTGF* expression (Chen et al. 2002; Holmes et al. 2001; Pickles & Leask 2007). The SBE is also critical for TGFB1-induced *CTGF* expression. Signalling through the type II TGFB receptor (TGFB2) results in activation of smad2 and smad3 nuclear factors. These factors in turn dimerize with smad4 to activate gene transcription through binding at the SBE. Smad3 and smad4, but not smad2 have been shown to be critical for TGFB1-induced

CTGF expression in fibroblasts, osteoblasts, and proximal-tubule epithelial cells (Holmes et al. 2001; Arnott et al. 2008; Phanish et al. 2006).

The SBE can also act in concert with the tandem Ets-1 sites and this synergy is required for the induction of *CTGF* by TGF β 1 in fibroblasts (Van Beek et al. 2006). These tandem Ets-1 binding sites have also been shown to be a target of the oncogenic transcriptional co-activators YAP and TEAD, and *CTGF* expression is potently activated by YAP in NIH-3T3 and MCF10A breast cancer cells (Zhao et al. 2008). An element termed the transcriptional enhancer dominant in chondrocytes (TRENDIC) can be activated by the matrix metalloprotease MMP3 (Eguchi et al. 2008). Transcriptional regulation is not a function commonly attributed to matrix metalloproteinases, however Eguchi and colleagues demonstrated nuclear localisation of MMP3 in chondrocytes and confirmed trans-activation of *CTGF* gene expression by nuclear MMP3 bound to the *CTGF* promoter (Eguchi et al. 2008). *CTGF* expression can also be modulated by hypoxia through two HIF response elements (HRE) at -3745 and -1558 (Higgins et al. 2004), however hypoxia has been shown to induce (Samarin et al. 2010) and inhibit (Kroening et al. 2009) *CTGF* expression depending on cell type and activation of different signalling pathways downstream from hypoxic stimuli, thus the role of hypoxia in regulating *CTGF* expression is controversial. Because *CTGF* is not normally expressed in lymphoid cells, it is unclear what transcriptional mechanisms are likely to be responsible for its aberrant expression in pre-B ALL.

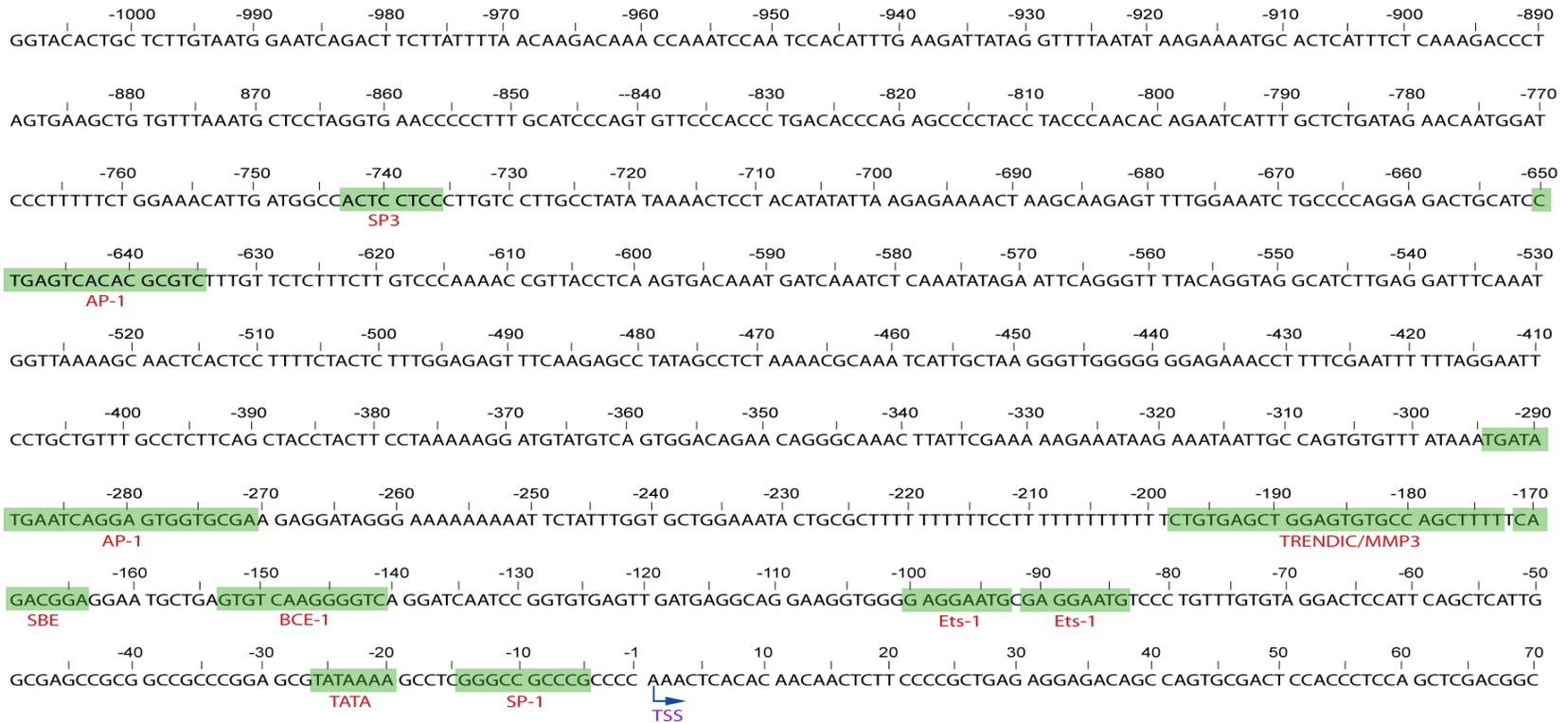


Figure 1.3 Transcription factor binding sites at the *CTGF* promoter

The *CTGF* promoter region from -1009 to +70 is shown. Functional transcription factor binding sites (green boxes) include two AP-1 sites (Yu et al. 2009), the transcriptional enhancer dominant in chondrocytes (TRENDIC) which is bound by MMP3 (Eguchi et al. 2007; Eguchi et al. 2008), an SP3 binding site (Fonseca et al. 2007), a smad binding element (SBE) (Arnott et al. 2008), basal control element (BCE-1) (Grotendorst 1997), tandem Ets-1 binding sites (Van Beek et al. 2006), a TATA box and an SP-1 binding site (Holmes et al. 2003). TSS: transcription start site (+1).

1.5.5.2 Epigenetic regulation of *CTGF* expression

Epigenetic features are now widely acknowledged to be important in tumorigenesis (Bonifer & Bowen 2010). Post-translational modifications of histone tails can alter chromatin structure, impacting on the accessibility of DNA elements for transcriptional initiation and elongation (Chi et al. 2010). Methylation of DNA at CpG dinucleotides can directly alter the ability of trans-factors to bind to DNA and can also recruit protein complexes that can remodel chromatin and promote gene inactivation, thus an increase in DNA methylation at gene promoters is generally associated with inactivation of gene transcription (Watanabe & Maekawa 2010). Cancer cells frequently exhibit global hypomethylation of DNA with an accompanying focal, hypermethylation at tumour suppressor loci (Kondo & Issa 2010). The role of epigenetic factors in the regulation of *CTGF* gene expression is still not clear, however several studies have identified cancer associated changes in DNA methylation and histone acetylation affecting the *CTGF* locus (Komorowsky et al. 2009; Chiba et al. 2004; Kikuchi et al. 2007). Chiba and colleagues identified an association between *CTGF* gene expression and histone acetylation, as well as changes in DNA methylation in some but not all hepatoma cell lines (Chiba et al. 2004; Chiba et al. 2005). An association between histone acetylation and *CTGF* expression has also been observed in renal endothelial cells (Komorowsky et al. 2009). DNA methylation of this region is inversely correlated with *CTGF* gene expression in ovarian cancer specimens and cell lines (Kikuchi et al. 2007). These findings suggest that aberrant epigenetic regulation may play a role in inducing *CTGF* expression in some cancers.

Recent evidence points to a crucial role for DNA methylation in haematological malignancies more generally. Milani and colleagues assessed the methylation status of 1320 CpG sites in the regulatory region of 416 genes and showed that the methylation status of 300 of these CpG sites facilitated the accurate prediction of lineage and cytogenetic subtype in a large cohort of paediatric ALL specimens (Milani et al. 2010). Furthermore, a recent analysis of the global DNA methylation status of 367 haematological neoplasms by Martin-Subero and colleagues identified increased

methylation of the *CTGF* locus in T-ALL and conversely a reduction in methylation of this locus in B-lineage ALLs (Martin-Subero et al. 2009). Given the notable absence of *CTGF* expression in T-ALL, this suggests that DNA methylation may be involved in regulating the *CTGF* locus in ALL (Martin-Subero et al. 2009). This study did however have significant limitations as only 1505 individual CpG sites were analysed, spread across 807 genes, and only 2 CpG residues were interrogated at the *CTGF* locus. Thus, a thorough investigation into epigenetic regulation of *CTGF* expression in ALL is warranted.

1.5.5.3 The *CTGF* 3' UTR

The 3' untranslated region (3' UTR) of canonical *CTGF* mRNA contains important regulatory elements and several miRNAs have been functionally demonstrated to regulate *CTGF* post-transcriptionally by inducing degradation of *CTGF* mRNA. These include miR18a, miR30, miR130 and the miR-17-92 cluster (Ohgawara et al. 2009; Duisters et al. 2009; Ernst et al. 2010). Another sequence element termed cis-acting element of structure-anchored repression (CAESAR) by Kubota and colleagues, has been shown to bind an as yet unidentified nuclear factor in chondrocytes, resulting in attenuation of CTGF protein synthesis (Kubota et al. 2000). Disruption of this sequence influences the stability of the *CTGF* mRNA, particularly in response to hypoxic conditions (Kubota et al. 2000; Kondo et al. 2006; Kubota et al. 2005). This element has the potential to act as a powerful inhibitor of CTGF protein expression and may serve as a negative regulator in other tissue types. While it is possible that modulation of *CTGF* mRNA stability conferred by the CAESAR element is mediated by miRNA, no functional miRNA binding domains have been identified in this 84bp sequence element.

1.5.6 *CTGF* expression and cancer

Aberrant expression of *CTGF* has been implicated in the biology of up to 21 different cancer types to date (Chu et al. 2008). However, the precise role that CTGF plays in tumour biology is presently unclear. This is highlighted by the observations that CTGF

can both promote and inhibit cancer progression, depending on the type and location of the tumour. *CTGF* expression has been associated with an aggressive phenotype in breast cancer (Kang et al. 2003), adenocarcinoma (Koliopoulos et al. 2002), pancreatic cancer (Wenger et al. 1999), malignant melanoma (Kubo et al. 1998) and glioblastoma (Yin et al. 2010). Conversely, *CTGF* expression has been associated with reduced metastasis and proliferation in ovarian cancer (Barbolina et al. 2009), chondrosarcoma (Shakunaga et al. 2000), lung cancer (Chien et al. 2006) and squamous cell carcinoma (Moritani et al. 2003). This duality likely reflects the importance of biological context and microenvironment on the action of *CTGF*.

CTGF expression has been shown to have prognostic relevance in ALL. A study investigating patient outcomes in adult ALL identified an association between high *CTGF* mRNA expression and a poor outcome (Sala-Torra et al. 2007). Patients were stratified based on the level of *CTGF* mRNA expression in their leukaemic blasts into three groups, low (n=26), intermediate (n=26) or high (n=26). The five year event free survival rates for these groups were 58%, 42% and 12% respectively (Sala-Torra et al. 2007). More recently, a COG study investigating gene expression classifiers for improved prognostic stratification of high-risk paediatric pre-B ALL patients identified an association between high *CTGF* expression and low relapse free survival (RFS) (Kang et al. 2010). Clearly, high *CTGF* expression is associated with a poor outcome in ALL and studies to establish the biological role of *CTGF* in this disease are warranted.

1.5.7 *CTGF* gene expression in the bone marrow

The bone marrow space is exposed to *CTGF* during normal development and maintenance, and *CTGF* is required for differentiation and proliferation of chondrocytes and osteoblasts (Takigawa et al. 2003; Nishida et al. 2000). The process of endochondral ossification whereby cartilaginous framework is replaced by bone, is distorted in the absence of *CTGF* protein (Song et al. 2007; Takigawa et al. 2003), and *CTGF* knock out mice die shortly after birth due to respiratory failure, as a result of significant skeletal

defects (Ivkovic et al. 2003). Other potential sources of CTGF protein in the bone marrow include fibroblasts, bone marrow stromal cells and endothelial cells.

Evidence supporting a role for *CTGF* expression in normal haemopoiesis has not been observed, however a 2007 study provided tantalizing data demonstrating that *CTGF* gene expression is induced in CD34^{pos} haemopoietic progenitor cells (HPCs) by adhesion to bone marrow stromal cells *in vitro* (Wagner et al. 2007). Adherent HPCs displayed 60-fold higher *CTGF* mRNA expression compared to non-adherent HPCs. Additional investigation is required to delineate the specific cell types expressing *CTGF* in this model system, as the CD34 marker can also be expressed by other cell types present in the bone marrow such as fibrocytes and vascular endothelial cells (Barth & Westhoff 2007; Nielsen & McNagny 2009). Nevertheless, this study points to a role for CTGF in adhesion of CD34^{pos} HPCs in the bone marrow and is the first such evidence of a role for CTGF in haemopoietic development.

1.5.8 CTGF, ALL and bone marrow fibrosis

High CTGF levels feature prominently in fibrotic pathologies and much research is directed toward understanding the role of CTGF in fibrotic disease (Wang et al. 2011; Brigstock 2010). In 1964 a study identified a concordance between bone marrow fibrosis and B-lineage ALL (Kundel et al. 1964), and this finding was further reinforced in 1989, with the reported observation that bone marrow fibrosis was present in 46 of the 56 (82%) paediatric ALL cases of B-lineage, but only 1 out of 7 patients with T-lineage ALL (Wallis & Reid 1989). Confirmation as to the importance of these early studies has been forthcoming with a report from Sweden by Noren-Nystrom and colleagues confirming increased bone marrow reticulin fiber density (RFD), a marker of fibrosis, in pre-B ALL versus T-ALL specimens (Noren-Nystrom et al. 2008). Furthermore, the clinical relevance of this increased bone marrow fibrosis was highlighted by the finding that the level of RFD at diagnosis correlated with the level of MRD after induction chemotherapy, suggesting that fibrotic bone marrow provides protection to leukaemic cells against chemotherapy (Noren-Nystrom et al. 2008). While this apparent fibrotic

pathology cannot be attributed retrospectively to the action of CTGF, the incidence of fibrosis and the predilection toward B-lineage ALL is consistent with the observed incidence of aberrant *CTGF* expression in pre-B ALL (~75%) (Boag et al. 2007; Vorwerk et al. 2000).

Solid tumours are now known to be comprised of not only neoplastic cells, but also stromal cells, including endothelial cells, inflammatory cells and fibroblasts (Chouaib et al. 2011; Allen & Louise Jones 2011). Cancer associated fibroblasts (CAFs) can enhance tumour growth and metastasis through release of classical growth factors, chemokines and enhanced synthesis of extracellular matrix (Kalluri & Zeisberg 2006). Elevated CTGF protein expression has been detected in the stroma of a number of tumours, including pancreatic, breast and oesophageal cancers (Wenger et al. 1999; Frazier & Grotendorst 1997; Koliopanos et al. 2002), suggesting an important role for CTGF in modulating the tumour microenvironment. In a xenograph model of human prostate cancer, human prostate carcinoma cells (LNCaP) were combined with murine fibroblasts carrying an inducible *CTGF* transgene. Expression of *CTGF* in murine fibroblasts significantly enhanced LNCaP engraftment and tumour growth, suggesting CTGF plays an important role in the tumour-promoting functions of CAFs (Yang et al. 2005).

While the role of CAFs in the growth and maintenance of solid tumours is now well established, the role of bone marrow stromal cells in promoting leukaemogenesis remains ill-defined. A number of stromal cell types exist in the bone marrow and these cells together with osteoblasts play important roles in regulating specialised HSC niches (Wilson & Trumpp 2006; Wilson et al. 2009). Furthermore, recent studies by Lee and colleagues have shown that CTGF can induce terminal differentiation of primary human bone marrow stromal cells into fibroblasts (Lee et al. 2010), thus ectopic expression of CTGF in the bone marrow microenvironment is likely to significantly alter the bone marrow stroma and may provide ALL cells with a growth advantage or refuge from chemotherapy (Lane et al. 2009).

1.6 Overall objectives of this study

Regulation of *CTGF* expression is complex and mechanisms vary significantly between cell types (Eguchi et al. 2007; Chaqour & Goppelt-Struebe 2006; Shi-Wen et al. 2008). While transcriptional networks have been elucidated in some cell types, the mechanisms by which *CTGF* expression is regulated in haemopoietic cells or leukaemia has not yet been investigated, as *CTGF* is not normally expressed in cells of haemopoietic origin. Transcription factors, epigenetic factors and miRNAs may all play a role in regulating *CTGF* expression and any number of these may be responsible for the observed dysregulation of the *CTGF* locus in pre-B ALL.

The specific aims of this thesis were as follows:

- i. To examine whether aberrant *CTGF* expression is a result of abnormal transcription factor expression or activation.
- ii. To examine clinical features associated with *CTGF* gene expression in pre-B ALL, and furthermore to investigate the association between *CTGF* expression and overall patient outcomes.
- iii. To determine if aberrant *CTGF* expression is a result of genomic lesions such as structural rearrangements or gene mutations.
- iv. To determine whether aberrant *CTGF* expression has an epigenetic basis.

Chapter 2

Materials and Methods

Chapter 2

Materials and Methods

2.1 Specimens and cell lines

2.1.1 Patient specimens

Bone marrow specimens were obtained from children diagnosed with ALL at Princess Margaret Hospital for Children, Perth, WA, Australia, between 1984 and 2005. Patients were treated according to Children's Cancer Group (CCG) protocols (Gaynon et al. 2000), and specimens were obtained either at the time of initial diagnosis, relapse or both. Ethics approval for this study was obtained from the Institutional Review Board, and informed consent for the use of tissues for research purposes was obtained for all individuals. White blood cells (WBC) were separated from whole BM specimens at the time of diagnosis using a Lymphoprep gradient (Nycomed) according to the manufacturer's instructions, then cryopreserved.

2.1.2 Donor mononuclear cells

Peripheral blood mononuclear cells (PBMCs) were isolated from healthy adult donors after provision of informed consent. Briefly, 8 ml of donor blood was collected into heparin vacutainers (BD Bioscience), inverted several times and left at room temperature for 15 min. Blood was separated by centrifugation at 1500 x g for 15 mins at 4°C. After centrifugation, the buffy coat was collected and adjusted to 10 ml with ice-cold PBS, washed twice with a further 10 ml of ice-cold PBS, then resuspended in one pellet volume of ice-cold PBS. PBMCs were processed immediately for nucleic acid extraction (described below), or stored at -80°C for no more than 2 weeks.

2.1.3 Mammalian tissue culture

The T-ALL cell line Jurkat was obtained from the American Type Tissue Collection (ATCC). All remaining ALL cell lines used in this study were generated at the Telethon

Institute for Child Health Research in the laboratory of Prof Ursula Kees according to previously published methods (Kees et al. 1987). Culture medium used for all cell culture was RPMI supplemented with 1 mM sodium pyruvate (MP Biomedicals, LLC), non-essential amino acids (MP Biomedicals, LLC), 2 nM 2-mercapto-ethanol, and 2 mM L-glutamine. Culture medium used for PER-145, PER-278, PER-371 and PER-377 was supplemented with 20% fetal calf serum (FCS) (Invitrogen, Life Technologies), while culture medium used for PER-485, PER-490, PER-495 and JURKAT was supplemented with 10% FCS. Cells were maintained in 5% CO₂ at 37°C.

2.1.3.1 Culture with 5-aza-2-deoxycytidine and Trichostatin A

Cells were seeded in fresh culture medium at 10⁶ cells/well in 24 well culture plates (Nunc, Thermo Fisher Scientific). In culture, 5-aza-2-deoxycytidine (Aza) was added to medium to a final concentration of 5 µM. Cells were treated for a total of 72 h with medium containing Aza replaced daily. Cells treated with Trichostatin A (TSA) were grown for 72 h, but treated for only the last 12 h in culture with TSA at a concentration of 330 nM. For cells treated with both agents, medium containing 5 µM Aza and 330 nM TSA was added to cells 60 h after the first Aza treatment.

2.2 Bioinformatics and statistics

2.2.1 Microarray data analysis

Global gene expression data from the PMH cohort was obtained by Boag and colleagues using the human Affymetrix U133A platform as described previously (Boag et al. 2007; Boag et al. 2006). Global gene expression data for the Ross cohort was obtained using the Affymetrix U133A platform according to the methods reported by Ross and colleagues (Ross et al. 2003). The Kang cohort was analysed using the Affymetrix plus 2.0 platform and global expression data was normalised according to methods outline in the report by Kang and colleagues (Kang et al. 2010). Normalised gene expression data for the Kang cohort was obtained from the European Bioinformatics Institute's Array

Express website <http://www.ebi.ac.uk/arrayexpress/> under the accession number E-GEOD-11877. Gene expression data for the Ross cohort was provided as supplemental data in the original publication (Ross et al. 2003).

2.2.1.1 Establishing gene expression cut-offs

Micro-array gene expression cut-offs were established based upon comparisons between global gene expression data from the PMH cohort and qRT-PCR validation experiments (Boag et al. 2007). Robust multi-array analysis (RMA) normalised probeset signals of < 50 were classed as negative for the mRNA targeted by that probeset. For *CTGF* mRNA expression levels, specimens were classed as high if the *CTGF* probeset intensity was > 500, med if between 500 and 50, and low/neg if < 50.

2.2.2 Promoter analysis

The MEME software suite (Bailey & Elkan 1994) was used to compare gene promoter regions for shared sequence motifs. Common motifs identified by analysis with MEME were screened for similarities with annotated transcription factor binding matrices contained in the JASPER database (Portales-Casamar et al. 2010) using the TOMTOM web interface (Gupta et al. 2007). Promoter regions were also manually scanned for transcription factor binding sites using the TFSEARCH database of transcription factor binding sites (Akiyama n.d.), and the PROMO transcription factor binding prediction software (Farre et al. 2003).

2.2.3 Statistical analyses

2.2.3.1 Spearman correlation analyses

Spearman's correlations were performed by Dr Martin Firth at the Telethon Institute for Child Health Research, Subiaco, Western Australia using the R software environment.

For each probeset contained within the three microarray datasets (PMH, Kang and Ross), its correlation with the *CTGF* probeset (209101_at) was calculated. This analysis was performed independently for all three microarray cohorts and probesets were ranked based upon correlation values. The top 100 *CTGF*-associated probesets were compared amongst cohorts.

2.2.3.2 Pearson correlation analyses

CTGF gene expression was measured in a selection of primary specimens from the PMH cohort by real-time PCR (section 2.3.4.1). These expression values were compared to those generated by microarray analysis of matched specimens in the PMH cohort described previously (Boag et al. 2007). Concordance was examined between these methods by carrying out a Pearson correlation analysis. This was performed using the “CORREL” function in the Excel spreadsheet application (Microsoft).

2.2.3.3 Graphing and statistical analysis of gene expression data

Graphing of all gene expression, patient survival and densitometry data was performed using Prism 4.0 (GraphPad Software). Statistical comparison of means was performed by t-tests or ANOVA in Prism 4.0.

2.3 Molecular Techniques

Unless otherwise stated, all biochemical and molecular reagents were obtained from Sigma Aldrich. A full list of suppliers is provided as an appendix at the end of this chapter in Table A2.1. A list of primers and oligonucleotides is presented as an appendix of this chapter in Table A2.2.

2.3.1 Isolation of nucleic acids

Prior to extraction of nucleic acids from bone marrow specimens or PMBCs, cryopreserved specimens were thawed rapidly at 37°C, transferred to ice, then processed according to extraction protocols below. Quantitation of nucleic acids was performed using an ND1000 spectrophotometer (ThermoFisher Scientific).

2.3.1.1 Genomic DNA extraction

After thawing of cryopreserved primary specimens or collection of cells from culture, 5×10^6 cells were washed twice in ice-cold PBS and then genomic DNA was extracted using the DNeasy Blood and Tissue kit (QIAGEN) as per the manufacturer's instructions. DNA was eluted in TE buffer and stored at 4°C until use.

2.3.1.2 Total RNA extraction

Total RNA was isolated from cells using a modified TRIZOL method (Hoffmann et al. 2005). After two washes with ice-cold PBS, cells were resuspended in one pellet volume of PBS and combined with 0.5 ml TRIZOL reagent (Invitrogen, Life Technologies) per 10^6 cells. Samples were vortexed thoroughly and left at room temperature for 5 min. After the addition of 0.1 ml chloroform, samples were vortexed thoroughly and centrifuged at 15,000 x g for 15 min at 4°C. The aqueous phase was then transferred to fresh nuclease-free, sterile 1.5 ml tubes and combined with 0.54 vol of absolute EtOH and vortexed. The resulting solution was purified using the RNeasy Mini kit (QIAGEN) following the manufacturer's instructions, from step 5 of the protocol: "Purification of total RNA from animal cells" in the RNeasy mini kit handbook. On-column DNase treatment was included during purification. Samples were eluted in 50 µl of nuclease free H₂O, quantitated and stored immediately at -80°C.

2.3.1.3 Cytoplasmic RNA extraction

Cytoplasmic RNA for northern blotting was isolated from ALL cell lines by the following method. Cells were collected by centrifugation at 3,000 x g then washed twice

with 10 ml of ice-cold PBS and resuspended in 200 μ l of lysis buffer [10mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM MgCl₂, 0.5% nonident P-40] per 5 x 10⁶ cells and incubated on ice for 10 min. Nuclei were collected by centrifugation at 12,800 x g for 10 min at 4°C. Supernatant containing cytoplasmic RNA was collected into fresh sterile, nuclease-free 1.5 ml tubes and combined with 25 μ l of 10% sodium dodecyl sulphate (SDS) on ice. RNA was extracted twice with 100 μ l of buffered phenol solution. After addition of the phenol solution, samples were vortexed thoroughly and centrifuged at 15,000 x g for 3 min at room temperature, and supernatants were transferred to fresh 1.5 ml tubes. Following the second phenol extraction, supernatants were combined with 0.1 vol of 2 M sodium acetate (pH 5.0) and 2.5 vol of absolute EtOH and incubated at -20°C overnight. RNA was pelleted by centrifugation at 15,000 x g for 15 min at 4°C. Pellets were washed with 70% EtOH, air dried briefly, and then resuspended in 50 μ l of nuclease-free H₂O and stored immediately at -80°C.

2.3.2 Polymerase chain reaction

Polymerase chain reaction (PCR) assays were performed on either a PTC 200 (MJ Research), or a C1000 Thermal Cycler (Bio-Rad). All PCR reactions were assembled on ice in sterile 0.2 ml thin-walled tubes (Bio-Rad). Primer and oligonucleotide sequences are listed in Table A2.

2.3.2.1 Standard polymerase chain reaction

Standard PCRs were performed using GoTaq Flexi DNA polymerase (Promega). Typically these reactions contained between 25 and 100 ng of template genomic DNA or equivalent cDNA, 5 μ l of 5 x reaction buffer, 1.5 mM MgCl₂, 200 μ M dNTPs, 0.2 μ M of each primer and 1 U of GoTaq DNA polymerase, combined into a 25 μ l reaction. DNA was denatured at 94°C for 2 min, followed by a variable number of cycles encompassing denaturation at 94°C for 30 s, primer annealing for 30 s (temperature determined using the melting temperature of primers), and a 30 s per kilobase extension

step at 72°C. The number of cycles varied depending on the abundance of the target as well as the efficiency of primer pairs.

2.3.2.2 High-fidelity polymerase chain reaction

When PCR was required to amplify cDNA for sequencing, or if long amplicons were generated, the LongAmp DNA polymerase (New England Biolabs) was used. The LongAmp polymerase exhibits 3'-5' exonuclease (proofreading) activity and thus ensured high-fidelity amplification. These reactions consisted up to 100 ng of template genomic DNA or equivalent cDNA, 5 µl of reaction buffer, 300 µM dNTPs, 0.4 µM of each primer, and 5 U of LongAmp DNA polymerase combined into a 25 µl reaction. Reactions were cycled as described for GoTaq above, however extension was performed at 65°C with a 1 min per kilobase extension time. For difficult to amplify templates or high GC-rich regions, 5% dimethyl sulfoxide (DMSO) was used in PCR reactions.

2.3.2.3 Agarose gel electrophoresis and gel extractions

PCR products were resolved by agarose gel electrophoresis (AGE). Molecular biology grade agarose (Amresco) was melted in 1 X TAE buffer [40 mM tris, 5.7% glacial acetic acid, 1 mM EDTA] and supplemented with SYBR Safe DNA stain (Invitrogen, Life Technologies). Samples were combined with 10 x loading buffer [4.4 mM Orange G, 40% glycerol] and separated by AGE alongside either 100 bp or 1 kb DNA markers (Invitrogen). DNA was visualised and photographed on a Gel Doc 2000 gel documentation system (Bio-Rad). PCR products destined for sequencing or for use as cDNA probes were isolated using the QIAquick gel extraction kit (QIAGEN) according to the manufacturer's protocol. Briefly, bands were excised using sterile scalpel blades, column purified and resuspended in 30 µl of H₂O. Extracts were stored at -20°C until required for downstream analysis.

2.3.2.4 First strand cDNA synthesis

First strand cDNA synthesis was performed using the Omniscript Reverse Transcriptase kit (QIAGEN) according to the manufacturer's instructions. RNA was first denatured at 65°C for 5 min to reduce secondary structure and then snap chilled on ice for a further 5 min. Reactions were assembled on ice and consisted of 1 µg of total RNA, 2 µl of reaction buffer, 500 µM dNTPs, 1 µM OligodT primer, 10 U of RNase inhibitor, and 4 U of Omniscript reverse transcriptase. Reactions were adjusted to 20 µl total volume with nuclease free H₂O mixed thoroughly and incubated at 37°C for 1 h. After extension, specimens were diluted 1 in 5 with nuclease free H₂O and stored at -80°C until required.

2.3.3 Rapid amplification of cDNA ends

Rapid amplification of cDNA ends (RACE) was performed using the GeneRacer kit (Invitrogen) to facilitate sequencing of full-length *CTGF* mRNA transcripts present in the pre-B ALL cell line PER-377. This method consists of three principal steps, preparing RNA for RACE amplification, reverse transcribing cDNA, and finally amplification by PCR.

2.3.3.1 Preparation of RACE-ready RNA

Ligation of a 44 base RNA oligonucleotide (5' GeneRacer Oligo) to the 5' end of decapped mRNA is the first step in the RACE protocol. The 5' GeneRacer oligo contains two priming sites that facilitate PCR amplification of the 5' end of transcripts in concert with a gene specific primer (GSP) after cDNA synthesis. Dephosphorylation and then decapping of 2 µg of DNase-treated total RNA was undertaken in accordance with the manufacturer's instructions without deviation. The 5' GeneRacer RNA oligonucleotide was then ligated to the 5' end of full length mRNA.

2.3.3.2 RACE cDNA synthesis

Full-length cDNA was reverse transcribed using SuperScript reverse transcriptase (RT) included in the GeneRacer kit and the GeneRacer OligodT primer which consists of 36 bases of complex sequence followed by 24 dT bases. This OligodT primer facilitated priming of first strand synthesis by annealing to polyadenylated mRNA and also incorporated known priming sites into the 3' end of all cDNA. Prior to reverse transcription, 2 µg of RACE-ready RNA, 0.5 mM dNTPs and 2.5 µM GeneRacer OligodT primer was combined in 3 µl H₂O and incubated at 65°C for 5 min to remove RNA secondary structure, then snap-chilled on ice for 2 min. The cDNA synthesis reaction consisted of 3 µl of denatured RACE-ready RNA, dNTPs and GeneRacer OligodT primer, combined in 20 µl with first strand reaction buffer, 5mM DTT, 40 U of RNase inhibitor and 200 U of SuperScript RT enzyme. Synthesis proceeded at 50°C for 1 h and this was followed by incubation at 70°C for 20 min to deactivate the RT enzyme. RNA was then digested by the addition of 2 U of RNase H at 37°C for 20 min.

2.3.3.3 Amplification of RACE clones by polymerase chain reaction

LongAmp DNA polymerase was used to amplify full-length race cDNA by nested PCR. For 5' RACE, the gene specific reverse primers used were CD_7R in the primary PCR, and CD_5R in the nested secondary PCR. For 3' RACE, the CD_1F forward primer was used in the primary PCR, followed by CD_1F_nested in the secondary PCR.

2.3.3.4 Band-stab polymerase chain reaction

The band stab technique (Bjourson & Cooper 1992), was used to isolate and re-amplify 5' and 3' RACE cDNA using LongAmp DNA polymerase. Bands of interest were inoculated into fresh PCR mix using sterile pipette tips and PCR amplification was performed using the same primer pair and thermocycling conditions as those used to amplify the original cDNA.

2.3.4 Quantitative real-time polymerase chain reaction

All reagents, assays, equipment and software were obtained from Applied Biosystems unless otherwise stated.

2.3.4.1 Quantitation of *CTGF* gene expression

Quantitative real-time PCR (qRT-PCR) was performed to measure expression of *CTGF* mRNA in cell lines and primary specimens. *CTGF* expression was measured using the on-demand assay Hs00170014_m1. Reactions were performed in 384 well plates on an ABI 7900HT Fast Real-Time PCR System, with the following thermal profile, 50°C for 2 min then 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. *CTGF* reactions were prepared by combining the following components per well, 1 µl of *CTGF* assay, 10 µl of universal PCR master mix, 8 µl of H₂O and 1 µl of first strand cDNA. *CTGF* gene expression was normalised using a custom *ACTB* assay comprising 10 µl of universal PCR master mix, 0.4 µl of 10 µM *ACTB* probe, 0.2 µl each of the *ACTB*_forward and *ACTB*_reverse primers (10 µM), 8.2 µl of H₂O and 1 µl of cDNA. Standards were prepared from cDNA generated from the PER-377 cell line. This cDNA was not diluted after cDNA synthesis and thus was approximately 5 times more concentrated than test specimens which were diluted 1 in 5 immediately after cDNA synthesis. Standards were prepared by serial dilution and assayed in duplicate to control for pipette error, while test specimens were assayed in triplicate. The relative expression for each assay was calculated using the standard curves.

2.3.4.2 *CTGF* gene copy number analysis

CTGF Gene copy number analysis was performed using qRT-PCR. The *CTGF* gene copy number assay Hs00517234_cn targets exon 5 within the coding region. This assay was run in duplex with the Taqman reference copy number assay for the *RPPHI* locus. Assays were performed in 384 well plates on an ABI 7900HT Fast Real-Time PCR System with the following thermal profile, 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. *CTGF* and *RPPHI* copy number assays were run in

duplex reactions prepared by combining the following components per well, 0.5 μ l *CTGF* assay, 0.5 μ l *RPPH1* reference assay, 5 μ l of universal PCR master mix, 2 μ l of H₂O and 2 μ l of genomic DNA at 5 ng/ μ l. Samples were assayed using 4 technical replicates. At the completion of the assay, data was imported into the Copy Caller software and analysed to determine *CTGF* gene copy number normalised to *RPPH1*.

2.3.5 Northern blotting of RNA

2.3.5.1 RNA electrophoresis and membrane transfer

Cytoplasmic RNA was isolated according to the method outlined in section 2.3.1.3, and 10 μ g of cytoplasmic RNA was combined with 1 μ l of 20 x northern running buffer (NRB) [122 mM Na₂HPO₄, 78 mM NaH₂PO₄] and 1.5 μ l of deionised 8.8 M glyoxal in a final volume of 20 μ l. Samples were denatured at 50°C for 1 h and then chilled on ice for 20 min. Samples were combined with 5 μ l of sample loading buffer [50% glycerol, 0.02% bromophenol blue, 1 x NRB] and together with an RNA ladder (Invitrogen, Life Technologies), were resolved on 0.8% agarose in 1 x NRB. Gels were run slowly at 50V, at 4°C for 3.5 h. RNA was stained with a solution of 0.8 mg/ml ethidium bromide in 1 x NRB for 30 min, then de-stained in fresh NRB for a further 30 min. RNA was transferred to a neutral magna membrane (Osmonics) by capillary transfer in 20 x SSC [3 M NaCl, 300 mM Na Citrate, pH 7.0] overnight and cross linked to the membrane while still damp with 150 mJoules UV using a GS Gene Linker (Bio-Rad).

2.3.5.2 Northern blot probe synthesis

Hybridisation probes were generated by PCR amplification using GoTaq DNA polymerase. The template cDNA used to amplify these probes was reverse transcribed from RNA isolated from an airway fibroblast cell line (HFL-1). This RNA sample was a kind gift from Mr Bradley Shelton, Lung Institute of Western Australia, (Shenton Park, Australia). Two probes were generated targeting *CTGF* mRNA. The first probe designated as the Central probe, hybridised *CTGF* between exons 3 and 5 and was

amplified with the primer pair NB_central_for and NB_central_rev. The second probe hybridized *CTGF* mRNA in the 3'UTR and was amplified with the primer pair NB_3pr_for and NB_3pr_rev. To provide a loading control, a probe hybridizing *ATP5G3* was amplified using the primer pair ATP5G3_for and ATP5G3_rev.

2.3.5.3 Radiolabelling of cDNA probes

After gel extraction, 30 ng of purified cDNA probe was labelled with 50 μ Ci of 32 P dCTP radionuclide (Perkin Elmer) using the Rediprime II Random prime labelling system (GE Healthcare) according to the manufacturers instructions.

2.3.5.4 Northern blot probe hybridisation

Directly after cross linking, the RNA-bound membrane was soaked in 2 x SSC for 5 min at room temperature, then placed between two sheets of nylon mesh and added to a 50 cm hybridisation bottle together with 15 ml of hybridization buffer [5% dextran sulphate, 20% Denharts Solution, 0.5 mg/ml salmon sperm DNA, 3% SSC, 0.1% SDS] pre-heated to 50°C. Hybridisation bottles were incubated at 50°C for 1 h in a model 400 hybridisation incubator (Robbins Scientific). 32 P-labelled probe was denatured by boiling for 5 min then snap chilled on ice for 10 min. Denatured probe was then added to the hybridisation bottle and incubated in the rolling incubator overnight at 50°C. The next morning, the membrane was washed with four successive and increasingly stringent 15 min washes, comprised of 2 x SSC, 2 x SSC/0.1% SDS, 0.5 x SSC/1% SDS, 0.2 x SSC/0.1% SDS respectively. The membrane was then incubated with autoradiograph film at -80°C. Membranes were stripped for re-probing by pouring a boiling solution of 0.1 x SSC/0.1% SDS over membranes and allowing them to cool to room temperature. Densitometry analysis was performed using the ImageJ program (NIH).

2.3.6 Southern blotting

2.3.6.1 Restriction digest of genomic DNA

Genomic DNA from pre-B ALL cell lines or PBMCs was concentrated by ethanol/sodium acetate precipitation. Briefly, DNA solutions were combined with 2.5 vol of absolute EtOH and 0.1 vol of 3 M sodium acetate pH 5.2 and stored overnight at -20°C. The following day, samples were centrifuged at 15,000 x g for 15 min at 4°C. Following removal of the supernatant, pellets were washed with 70% EtOH and briefly air-dried. Once almost dry, pellets were resuspended in 50 µl nuclease-free H₂O. Samples were heated at 65°C for 30 min to ensure complete resuspension, and then lightly vortexed to avoid excessive pipetting which may result in shearing of full-length DNA. Samples were quantitated and 15 µg of gDNA was digested with 15 U of either Psi I (New England Biolabs) or Hind III (Promega) in 150 µl digests overnight at 37°C. The following day, restriction enzymes were inactivated by incubating samples at 65°C for 15 min and samples were concentrated by EtOH/sodium acetate precipitation as described above and resuspended in 15 µl of nuclease-free H₂O.

2.3.6.2 Electrophoresis and transfer of digested genomic DNA

Samples were separated on 0.8% agarose in TAE overnight at 40 V. The following day, the gel was stained with 1 x SYBR safe DNA stain (Invitrogen, Life technologies) in TAE buffer for 30 min and photographed under UV illumination to confirm adequate migration of digested DNA. The gel was then soaked in 300 ml of denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 40 min at room temperature, then rinsed twice with ddH₂O (Baxter) and soaked in two changes of neutralizing solution (0.5 M tris pH 7.4, 1.5 M NaCl) for a total of 1 h. DNA was then transferred to a neutral Magna membrane (Osmonics) by capillary transfer with 10 x SSC overnight. Following transfer, lanes were marked with a soft pencil and the membrane was soaked in 6 x SSC for 5 min. DNA was cross linked to the membrane while still damp with 150 mJoules UV using a GS Gene Linker (Bio-Rad).

2.3.6.3 Southern blot probe synthesis

A hybridisation probe for genomic Southern blotting was amplified from PBMC genomic DNA using LongAmp DNA polymerase. This DNA probe spanned the *CTGF* gene from -1024 to +3418 and was amplified with the primer pair SB_CTGF_for and SB_CTGF_rev. A 1050 bp probe which hybridized the *IL2RB* gene was a kind gift from Dr M. Hatakeyama, Osaka (Hatakeyama et al. 1989).

2.3.6.4 Radiolabelling of probes

After gel extraction, 30 ng of gel purified probe was labelled with 50 μ Ci of 32 P dCTP radionuclide (Perkin Elmer) using the Rediprime II Random prime labelling system (GE Healthcare) according to the manufacturers instructions.

2.3.6.5 Southern probe hybridisation

Membrane and mesh was pre-soaked in 2 x SSC for 5 min at room temp, then rolled and added to a hybridisation bottle together with 15 ml of pre-heated hybridization buffer [2 x SSPE, 7% SDS, 0.5% Blotto, 0.5 mg/ml salmon sperm DNA] and incubated at 65°C for 1 h. 32 P-labelled DNA probe was denatured by boiling for 5 min then snap chilled on ice for 10 min. Denatured probe was then added to the hybridisation bottle and incubated overnight at 65°C. The next morning, the membrane was washed with four successive 15 min washes comprised of 2 x SSC/0.1% SDS, 2 x SSC/0.1% SDS, 1 x SSC/0.1% SDS, 0.5 x SSC/0.1% SDS. The membrane was then incubated with autoradiograph film at -80°C. Membranes were stripped for re-probing by pouring a boiling solution of 0.1 x SSC/0.1% SDS over membranes and allowing them to cool to room temperature.

2.3.6.6 Electrophoresis and transfer of RACE-PCR products

Band-stab PCR products were examined by Southern blotting with a *CTGF*-specific probe. PCR products were separated on 1.5% agarose in TAE and stained with SYBR

Safe DNA stain. PCR products were then transferred to a positively charged N⁺ Hybond membrane (Amersham, GE Healthcare) by capillary transfer with 0.4 M NaOH overnight. Following transfer, lanes were marked with a soft pencil and the damp membrane was cross linked with 150 mJoules UV using a GS Gene Linker (Bio-Rad). The probe used to hybridise band-stab PCR products was amplified from cloned full-length *CTGF* cDNA with GoTaq DNA polymerase using the CD_1F and CD_5R primers, then radiolabelled and hybridised to membrane-bound RACE PCR products according to the method outlined above in section 2.3.6.5.

2.3.7 DNA cloning

2.3.7.1 Preparation of chemically competent cells

TOP10 *E. coli* were a gift from Mrs Marie Scobie. Single colony isolation was performed by streak isolating on LB agar plates grown at 37°C overnight. A 10 ml culture of LB broth [10 g Bacto-tryptone (BD Bioscience), 5 g yeast extract (BD Bioscience), 170 mM NaCl] was inoculated with a single colony and grown overnight at 30°C with moderate shaking. The following day, 99 ml of LB agar was pre-warmed to 30°C, and then inoculated with 1 ml of overnight culture. This culture was grown at 30°C with moderate shaking and the absorbency at 600 nm was measured regularly using a BioPhotometer spectrophotometer (Eppendorf). Once the OD₆₀₀ reached 0.5, cultures were immediately transferred to a 4°C cold-room and chilled on ice for 10 min. All remaining steps were where possible performed at 4°C using pre-chilled glassware and instruments. Cells were collected by centrifugation at 3000 x g for 10 min. The supernatant was discarded and cells were gently resuspended in 20 ml of ice-cold, 50 mM CaCl₂. Cells were collected by centrifugation at 3000 x g for 5 min and then resuspended gently in ice-cold 75 mM CaCl₂/25% glycerol. This suspension was added to pre-chilled, sterile 1.5 ml tubes in 100 µl aliquots, then snap frozen by immersion in liquid nitrogen. Competent cells were stored at -80°C until use.

2.3.7.2 Cloning into the pGEM-T Easy vector

PCR products were cloned using the pGEM-T Easy TA cloning system (Promega) according to the manufacturer's protocol. Inserts were invariably generated using the LongAmp DNA polymerase. While this enzyme exhibits 3'-5' exonuclease activity, the majority of PCR products contain dA overhangs at the 3' end allowing cloning into TA vectors such as pGEM-T Easy. After gel purification, PCR products were quantitated and 5 to 10 ng of insert was combined with ligation buffer, pGEM-T Easy vector (50 ng) and 3 U of T4 DNA ligase in a 10 µl reaction. Ligations were incubated overnight at 4°C.

2.3.7.3 Transformation of chemically competent cells

Chemically competent cells were thawed on ice for 15 min. After thawing, 5 µl of ligation reaction was added to competent cells followed by a further 15 min incubation on ice. Cells were then subjected to heat shock transformation by submersion in a 42°C water bath for 1 min, then transferred to ice immediately and incubated for 15 min. To maximise transformation, this process of heat shock and recovery was repeated once more. After the second 15 min recovery, 250 µl of SOC medium [10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose] was added to cells and tubes were incubated for 1 h at 37°C with moderate shaking. LB agar plates were prepared containing 100 µg/ml ampicillin. Thirty mins prior to spreading transformants, plates were supplemented with 50 µl of 200 mM IPTG and 20 µl of 50 mg/ml X-Gal and placed at 37°C. At the completion of the 1 h incubation, 150 µl of transformed cells were spread on the prepared selective agar plates and incubated overnight at 37°C.

2.3.7.4 Transformant screening and plasmid DNA extraction

The following day, blue-white screening revealed clones carrying DNA inserts and these clones were used to inoculate LB broth supplemented with 100 µg/ml ampicillin. These 5 ml cultures were grown overnight at 37°C with vigorous shaking. The following day, glycerol stocks were prepared from each culture by combining 100 µl of culture with

100 µl of 80% glycerol. These stocks were frozen immediately at -80°C. The remainder of the overnight culture was centrifuged in preparation for plasmid DNA extraction using the QIAprep Spin Miniprep kit (QIAGEN) according to the manufacturer's instructions. Purified plasmid DNA was eluted in 50 µl of elution buffer and stored at -20°C.

2.3.7.5 Isolation of bacterial artificial chromosome DNA

The Human TilePath Library clone RP11-69I8 was purchased from the Murdoch Research Institute (Melbourne, Australia). This bacterial artificial chromosome (BAC) contained a fragment of chromosome 6 that included the *CTGF* locus. A single colony was streak isolated on an LB agar plate supplemented with 12.5 µg/ml chloramphenicol. This colony was used to inoculate 50 ml LB broth containing the same concentration of chloramphenicol and this culture was grown overnight at 37°C with shaking. The following day, cells were collected by centrifugation at 3000 x g for 10 min at 4°C. Cells were resuspended in 50 ml ice-cold STE buffer [10 mM Tris-Cl (pH 8.0), 0.1 M NaCl, 1 mM EDTA (pH 8.0)]. Cell were then collected by centrifugation again and resuspended in 1 ml of ice-cold alkaline lysis solution I [50 mM glucose, 25 mM Tris-Cl (pH 8.0), 10 mM EDTA (pH 8.0)] and transferred to five 1.5 ml tubes (200 µl each) and placed on ice. After 5 min, 400 µl of alkaline lysis buffer II [0.2 M NaOH, 1% w/v SDS] was added to tubes and mixed by gentle inversion several times. Next, 300 µl of ice-cold alkaline lysis buffer III [3 M potassium, 5 M acetate] was added and tubes were mixed gently by inversion and incubated on ice for 5 min. Debris was cleared by centrifugation at 15,000 x g for 5 min at 4°C and supernatants were decanted into fresh 2 ml tubes. At room temperature, 900 µl of isopropanol was added and the tube was mixed again by gentle inversion. Precipitated nucleic acids were then collected by centrifugation at 15,000 x g for 5 min at room temperature. Pellets were washed once with 70% EtOH and briefly air dried, then resuspended in 20 µl of TE buffer. Tubes were combined and stored at -20°C.

2.3.8 DNA Sequencing

DNA sequencing was carried out using BigDye version 3.1 (Applied Biosystems). Sequencing reactions were performed in 96 well plates suitable for the 3130x Genetic Analyser sequencing platform (Applied Biosystems). Briefly, 75 ng of plasmid DNA or 100 ng of gel extracted PCR products were combined with 3.5 μ l of 5 x BigDye reaction buffer, 9.3 μ l of ultrapure H₂O (Baxter), 3.2 μ l of 1 μ M sequencing primer and 1 μ l of BigDye v3.1. Samples were denatured at 96°C for 1 min, followed by 25 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 4 min. Inserts cloned into the pGEM-T Easy vector were sequenced using M13 forward and reverse primers listed in Table A2.

2.3.8.1 BigDye reaction purification and analysis

BigDye reaction products were purified *in situ* by EtOH/sodium acetate precipitation. To each well, 2 μ l of 3 M sodium acetate (pH 5.2) and 50 μ l of absolute EtOH was added. The plate was resealed, mixed by vortexing, and spun for 30 s at 1000 x g to collect the solution to the bottom of wells. The plate was then incubated at room temperature in the dark for 2 h. BigDye products were then pelleted by centrifugation at 15,000 x g for 30 min at 4°C. The supernatant was removed by inversion and pellets were washed using 70% EtOH and centrifuged again at 15,000 x g for 10 min at 4°C. Supernatants were removed by inversion and the pellets were dried by inverting the plate on lint-free paper towel and centrifuging at 800 x g for 1 min.

After briefly air drying pellets, 10 μ l of HiDi formamide (Applied Biosystems) was then added to each well and the pellets were denatured by two consecutive rounds of 96°C for 2 min followed by incubation on ice for 2 min. Plates were then loaded onto a 3130x Genetic Analyzer (Applied Biosystems) for capillary sequencing. Sequencing data generated from BigDye sequencing was analysed using the KB Base Caller software (Applied Biosystems).

2.3.8.2 *CTGF* promoter and 3'UTR sequencing

The promoter and 3' UTR regions of the *CTGF* locus were amplified using the high-fidelity LongAmp DNA polymerase. The PR_3F and PR_1R primers were used to generate a PCR amplicon spanning the *CTGF* proximal promoter and this product was gel extracted and sequenced using the primers; PR_3F, PR_3R, PR_2F, PR_2R, PR_1F and PR_1R. The 3'UTR was amplified in 3 separate PCR amplicons using primer pairs; CD_7F and CD_7R, CD_8F and CD_8R, CD_9F and CD_9R and these three amplicons were gel purified and sequenced in forward and reverse orientation using these primers.

2.3.9 Analysis of DNA methylation

2.3.9.1 Conversion of DNA by sodium bisulfite

To investigate DNA methylation at the *CTGF* locus a number of techniques were performed that relied upon effective bisulfite conversion of non-methylated cytosines in the genome by sodium bisulfite. The EpiTect Bisulfite Kit (QIAGEN) was used according to the manufacturer's instructions to convert genomic DNA from cell lines and primary specimens, as well as BAC DNA which served as a control in methylation specific PCRs (outlined below).

2.3.9.2 BAC DNA methylation

The human BAC Tile Path clone RP11-69I8 containing the entire *CTGF* coding region was methylated using the CpG methyltransferase M.SssI (New England Biolabs) to serve as a positive control in methylation specific PCRs. Reactions consisted of 1 µg of BAC DNA, 160 µM S-adenosylmethionine (SAM), reaction buffer and 4 U of methyltransferase in 20 µl. Reactions were performed at 37°C for 3 h. The enzyme was then deactivated by incubating at 65°C for 20 min. Converted BAC DNA was diluted to 1000 µl with H₂O and stored at -20°C.

2.3.9.3 Methylation specific PCR

Bisulfite converted DNA was subjected to PCR amplification using primers specific for either methylated or unmethylated priming sites. These primer pairs were designated “methylated-site PCR” (MSP_for, MSP_rev) or “unmethylated site PCR” (USP_for, USP_rev), and have been used successfully in other studies (Chiba et al. 2005). Methylation specific PCR (msPCR) reactions were performed with the GoTaq DNA polymerase as outlined in section 2.3.2.1, using 100 ng of bisulfite-treated genomic DNA. Bisulfite-treated BAC DNA (1 ng) served as a positive control for the USP reaction, while M.SssI-methylated, bisulfite-treated BAC DNA served as a positive control for the MSP reaction.

2.3.9.4 Bisulfite Sequencing

Three separate PCR amplicons spanning the *CTGF* CpG island were amplified using LongAmp DNA polymerase. These amplicons, described previously (Kikuchi et al. 2007) were generated from 100 ng of bisulfite treated genomic DNA. Amplicon 1 was amplified with the primer pair Bis1_for and Bis1_rev. Bisulfite amplicons 2 and 3 were amplified in nested PCR reactions using the following primers; Bis2-primary_for and Bis2_primary_rev, Bis2_secondary_for and Bis2_secondary_rev, Bis3-primary_for and Bis3_primary_rev, Bis3_secondary_for and Bis3_secondary_rev. PCR products were gel extracted using the QIAquick Gel Extraction Kit (QIAGEN) and eluted in 50 µl of H₂O. Purified bisulfite amplicons were cloned into the pGEM-T Easy vector according to the method outlined in section 2.3.7.2 and plasmid DNA was purified for sequencing from a selection of transformants.

Sequencing of cloned bisulfite amplicons was performed according to the method described in section 2.3.8, however bisulfite-converted DNA has a significantly lower melting temperature, and thus sequencing of bisulfite clones required a different thermal profile for BigDye reactions. After denaturation at 96°C for 1 min, samples were subjected to 24 cycles of 96°C for 10 s followed by extension at 50°C for 4 min. Bisulfite clones were analysed for effective bisulfite conversion and incidence of CpG methylation using BiQ Analyzer software (Bock et al. 2005).

2.3.10 Immunological detection of CTGF protein

2.3.10.1 Isolation of total cellular protein

Total protein was isolated from cell lines for analysis by western blotting. Cells were pelleted by centrifugation and washed twice with ice-cold PBS. Cell pellets were resuspended in one pellet volume of ice-cold PBS and then lysed using 50 μ l of ice-cold protein lysis buffer [2 mM ethylenediaminetetraacetic acid (EDTA), 10 mM iodoacetamide, 25 μ g/ml p-nitrophenylguanidinobenzoate, 10 μ g/ml Leupeptin, 10 μ g/ml Aprotinin and 0.5% Triton-X] per 10^6 cells, and incubated on ice for 30 min. Samples were centrifuged for 20 min at 15,000 x g at 4°C and supernatants were transferred to fresh 1.5 ml tubes. Protein samples were stored at -20°C or used immediately in downstream assays.

2.3.10.2 Protein estimation

Protein concentration in cell line extracts was determined using a commercial Bradford assay (Bradford 1976). Protein concentration was re-established each time samples were thawed prior to western blotting. Bio-Rad Protein Assay (Bio-Rad) dye was diluted 1 in 4 with H₂O and then added to a 96 well micro-titre plate (Becton Dickinson) at 200 μ l per well. A stock solution of 10 mg/ml BSA was serially diluted to provide a standard curve within the range of 5 mg/ml to 10 μ g/ml. Samples were diluted 1 in 10 and together with standards, were added to the plate at 10 μ l per well. Standards were assayed in duplicate, while protein samples were assayed in triplicate. After addition of samples and standards, the plate was incubated at room temperature for 5 min, and the absorbance was measured at 595 nm on a Victor² 1420 multilabel counter (Wallac, Perkin Elmer). Protein concentration in cell line extracts was calculated using the standard curve.

2.3.10.3 Sample preparation

Prior to SDS polyacrylamide gel-electrophoresis (SDS-PAGE), 75 µg of total protein was combined with 3.3 x sample buffer [0.2 M Tris-HCl, 40% glycerol, 8.3 mM dithiothreitol, 4% w/v SDS, 0.008% bromophenol blue and 6% 2-mercaptoethanol] and placed in boiling water for 5 min. Samples were chilled on ice for 5 min then loaded onto the gel or frozen immediately at -20°C until required. If frozen, samples were denatured by boiling again prior to electrophoresis.

2.3.10.4 Heparin-affinity enrichment of conditioned culture medium

Cells were seeded in fresh culture medium at a density of 10^6 cells per ml on 24 well culture plates (Nunc, Thermo Fisher Scientific), and grown for 24 h prior to isolation of conditioned medium (CM). Cells and CM were collected and centrifuged at 800 x g for 5 min and supernatants comprising the CM was then transferred to fresh tubes and chilled on ice. After 5 min, 5 ml of ice-cold PBS was added to fresh 10 ml tubes together with 150 µl of heparin agarose beads. 5 ml of chilled CM was added to these tubes and incubated for 24 h at 4°C with constant gentle mixing. After 24 h, beads were collected by centrifugation at 800 x g for 5 min at 4°C then allowed to stand for 2 min at 4°C. Supernatant was carefully removed by vacuum aspiration and beads were washed twice by this method with ice-cold PBS. After the second wash, beads were resuspended in 1 ml ice-cold PBS and transferred to 1.5 ml tubes. Beads were collected by centrifugation at 800 x g for 3 min at 4°C and the supernatant was carefully aspirated. Bound protein was eluted by adding 200 µl of protein loading buffer and denatured in boiling water for 5 min. Tubes were centrifuged while still hot at 10,800 x g for 3 min to collect beads. The supernatant containing eluted protein was removed to fresh 1.5 ml tubes and either subjected immediately to SDS-PAGE or frozen at -80°C until required.

2.3.10.5 Western blotting

Denatured protein samples were separated by 15% SDS-PAGE together with Precision Plus Protein Dual Colour Standards (Bio-Rad), before being transferred to Hybond-C

Super membrane (AP Biotech) in transfer buffer [25 mM tris, 190 mM glycine, 20% methanol, pH 8.2]. Membranes were blocked with 5% skim milk powder (SMP) in wash buffer [50 mM tris, 150 mM NaCl, 0.05% Tween] at room temperature for 1 h and incubated with primary antibody in 5% SMP in wash buffer overnight at 4°C. Primary antibody was removed and the membrane rinsed twice with wash buffer, then washed three times for 15 min with wash buffer. Membranes were incubated with either horse radish peroxidase (HRP) or biotin conjugated secondary antibodies, in 5% SMP in wash buffer for 1 h at room temperature. After three 15 min washes with wash buffer, membranes incubated with biotinylated secondary antibodies were incubated with a streptavidin conjugated HRP (Dako) for 5 min then re-washed three times in wash buffer. Visualisation of bands was performed by enhanced chemiluminescence using the Immobilon HRP substrate (Millipore).

2.3.10.6 Antibodies

The two CTGF antibodies used in this study included the polyclonal L20 goat anti-CTGF antibody (Santa Cruz Biotechnology), which targets amino acids 150-200 of the CTGF protein, and the polyclonal Rabbit-A rabbit anti-CTGF antibody, which targets amino acids 247-260 of the CTGF protein. This antibody was a generous gift from Prof David Brigstock (Brigstock et al. 1997). The ACTN05 monoclonal mouse anti-ACTIN antibody (Neomarkers) was used to detect ACTIN which served as a loading control. Secondary, HRP-conjugated antibodies included the anti-rabbit Na9340V antibody (GE Healthcare) and the anti-mouse Na 9310V antibody (GE Healthcare). For detection of primary goat antibodies, the biotinylated EO466 rabbit anti-goat antibody (Dako) was used followed by detection with streptavidin-conjugated HRP (Dako).

2.3.10.7 Membrane stripping protocol

Where it was necessary to strip and re-probe membranes, they were incubated in stripping buffer [50 mM Tris (pH 6.8), 2% SDS, 100 mM 2-mercaptoethanol] at 50°C for 30 min with gently shaking, then washed three times in western wash buffer for 15

min at room temperature. These membranes were then blocked and probed as previously described.

Table A2.1 List of suppliers

Amresco	- Solon, Ohio, USA
AP Biotech	- Little Chalfont, Buckinghamshire, UK
Baxter	- Old Toongabbie, New South Wales, Australia
BD Bioscience	- San Jose, California, USA
Becton Dickinson	- Franklin Lakes, New Jersey, USA
Bio-Rad	- Hercules, California, USA
Dako	- DK-2600, Glostrup, Denmark
GE Healthcare	- Little Chalfont, Buckinghamshire, UK
Geospiza	- Seattle, Washington, USA
GraphPad Software	- De La Playa, California, USA.
Ibis Biosciences	- Carlsbad, California, USA
Life Technologies	- Carlsbad, California, USA
Microsoft	- Redmond, Washington, USA
Milipore	- Bedford, Massachusetts, USA
MJ Research	- Waltham, Massachusetts, USA
MP Biomedicals, LLC	- Santa Ana, California, USA
Neomarkers	- Fremont, California, USA
New England Biolabs	- Ipswich, Massachusetts, USA
Nycomed	- Asker, Oslo, Norway
Osmonics	- Westborough, Massachusetts, USA
Perkin Elmer	- Waltham, Massachusetts, USA
QIAGEN	- Hilden, Germany
Robbins Scientific	- Sunny Vale, California, USA
Santa Cruz Biotechnology	- Santa Cruz, California, USA
Sigma Aldrich	- St Louis, Missouri, USA
Thermo Fisher Scientific	- Waltham, Massachusetts, USA

Table A2.2 Primer and oligonucleotide sequences

<u>primer name</u>	<u>sequence 5'-3'</u>
ACTB_forward	GGCACCCAGCACAATGAAG
ACTB_probe	VIC_TCAAGATCATTGCTCCTCCTGAGCGC_TAMRA
ACTB_reverse	GCCGATCCACACGGAGTACT
ATP5G3_for	TAATCCAAAGGGAGTTTCAGAC
ATP5G3_rev	AATCAAGAAAGCAACCATCAAAC
Bis1_for	GTAGGAAGGTGGGGAGGAA
Bis1_rev	CACTAACTATCTCCTCTCAAC
Bis2_primary_for	GGAATGTTGAGTGTTAAGGGGTTAGGATTA
Bis2_primary_rev	ATCAAACATTA AAAACTCTCACATCCAAA
Bis2_secondary_for	TTGAGAGGAGATAGTTAGTG
Bis2_secondary_rev	AACAAAATAAACCCCTTATAC
Bis3_primary_for	GGTTGTTAGGGAGGGATT
Bis3_primary_rev	TCCATACTACACAAAACATACAACC
Bis3_secondary_for	GTATAAGGGTTTATTTTGTATTT
Bis3_secondary_rev	CACTAATACTTACA ACTACTCTA
CD_1F	AGTGCGACTCCACCCTCCA
CD_1F_nested	CCTCCAGCTCGACGGCAG
CD_4F	AAGGACCAAACCGTGGTTGG
CD_5F	GACCTGGAAGAGAACATTAAGG
CD_5R	TACTCCACAGAATTTAGCTCGG
CD_6R	TTAATGTCTCTCACTCTCTGGC
CD_7F	TGTACTACAGGAAGATGTACGG
CD_7R	AGAATGTCAGAGCTGAGTCTGC
CD_8F	TTAGCGTGCTCACTGACCTG
CD_8R	TTCTACCTAGAAATCAGCC
CD_9F	TGCTCAGATAGAATGACAGTCC
CD_9R	TCAATGTGGTGTGTATGCCTGC
GeneRacer 3'	GCTGTCAACGATACGCTACGTAACG
GeneRacer 3' nested	CGCTACGTAACGGCATGACAGTG

Table A2.2 (continued)

<u>primer name</u>	<u>sequence 5'-3'</u>
GeneRacer 5'	CGACTGGAGCACGAGGACTGA
GeneRacer 5' nested	GGACTGACATGGACTGAAGGAGTA
GeneRacer OligodT	GCTGTCAACGATACGCTACGTAACGGCATGACAGTGT(24)
GeneRacer RNA oligo	CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAGUAGAAA
M13_for	GTAAAACGACGGCCAGTG
M13_rev	CAGGAAACAGCTATGAC
MSP_for	TCGTTTCGGTCGATAGTTTC
MSP_rev	CGAAACCCATACTAACGACG
NB_3pr_for	AGGGTACCAGCAGAAAGGTTAGTA
NB_3pr_rev	AGAAATCAGCCTGCCAAGGACT
NB_central_for	GAGTGGGTGTGTGACGAGCCCAAGG
NB_central_rev	ATGTCTCCGTACATCTTCCTGTAGT
PR_1F	AGTGGACAGAACAGGGCA
PR_1R	GCGGCTGCCGTCGAGCTG
PR_2F	CAGGTAGGCATCTTGAG
PR_2R	CACTGGCTGTCTCCTC
PR_3F	ACAACAGGGTACACTGCTC
PR_3R	TGATTTGCGTTTTAGAGGC
SB_CTGF_for	CCTTTGCTGGCCATTCACACTATTG
SB_CTGF_rev	ACCCCTACTAGATTCAACAGCTGCT
USP_for	TTGTTTTGGTTGATAGTTTT
USP_rev	CAAAACCCATACTAACAACA

Chapter 3

Gene Expression and Clinical Correlates of *CTGF* Expression in Childhood Pre-B ALL

Chapter 3

Gene Expression and Clinical Correlates of *CTGF* Expression in Childhood Pre-B ALL

3.1 Introduction

Oncogenic events contributing to leukaemogenesis include the activation of proto-oncogenes that promote dysregulation of cell cycle controls, apoptosis, and differentiation (Pui 2006). Aberrant *CTGF* gene expression occurs in around 75% of pre-B ALL cases (Vorwerk et al. 2000; Boag et al. 2007; Sala-Torra et al. 2007), however the mechanism underlying this dysregulated expression is not known. Activation of *CTGF* transcription in other cell types has been attributed to numerous soluble factors such as TGFB1 (Arnott et al. 2008), SRF (Muehlich et al. 2007) and MMP3 (Eguchi et al. 2008), as well as changes in the microenvironment such as hypoxia (Hong et al. 2006), mechanical stretch (Nishida et al. 2008) and high glucose (Wang et al. 2008). Regulation of the *CTGF* promoter is complex and is heavily dependent on cell type and tissue context (Cicha & Goppelt-Struebe 2009). Control of *CTGF* expression and its role in the cellular biology of lymphocyte precursors remains ill-defined as there has only been one study to date that has identified *CTGF* gene expression in normal haemopoietic precursor cells (Wagner et al. 2007), and no detailed characterisation of *CTGF* gene expression in haemopoietic cells has been reported.

Analysis of global gene expression data may provide some clues. Identifying transcriptional programs associated with *CTGF* gene expression by discerning genes that are highly correlated with *CTGF*, could point to the involvement of common transcription factors or regulatory pathways (Eisen et al. 1998). Global gene expression data gathered from a cohort of 73 paediatric patients diagnosed with pre-B ALL has been recently generated by this laboratory (Boag et al. 2006; Boag et al. 2007). This has provided a valuable resource to investigate whether any genes display patterns of expression similar to *CTGF*. This approach was expected to shed light on putative DNA

regulatory elements or transcription factors contributing to *CTGF* deregulation, through analysis of the regulatory sequences of co-regulated genes in pre-B ALL.

While its role in the leukaemogenesis of pre-B ALL is unclear, several studies have highlighted associations between *CTGF* gene expression and patient outcome in high-risk pre-B ALL. Sala-Torra and colleagues demonstrated a significant difference in EFS in adult ALL patients stratified based upon *CTGF* mRNA expression levels exhibited by their leukaemic cells. Specifically, patients whose leukaemic blasts expressed *CTGF* at low levels showed 5yr EFS of 58%, in contrast to 12% for those with high *CTGF* expression (Sala-Torra et al. 2007). Clearly, high *CTGF* expression was associated with a profoundly poor outcome for these patients. In relation to childhood ALL, a recent COG study investigating methods of improving stratification of high-risk pre-B ALL patients by microarray, successfully used *CTGF* expression in a multi-gene classifier model to predict the response of patients to therapy (Kang et al. 2010). As in adult ALL, high *CTGF* expression in this paediatric cohort was associated with a poor outcome and shorter period of RFS. These studies support the hypothesis that *CTGF* protein may play an important role in the pathobiology of ALL. *CTGF* expression has been associated with a metastatic and aggressive phenotype in other cancers (Shimo et al. 2006; Aikawa et al. 2006; Kang et al. 2003), however despite clear association with patient outcomes, no progress has been made in identifying biological features associated with *CTGF* gene expression in pre-B ALL. Analysis of clinical features associated with *CTGF* expression may point to a role for this pleiotropic protein in leukaemogenesis or maintenance of leukaemic cells.

In this study, analysis of *CTGF* gene expression in primary paediatric pre-B ALL specimens was performed to establish gene expression and clinical correlates with *CTGF*. Specifically, the aims of this chapter were to 1) validate *CTGF* expression data generated by microarray in primary specimens using qRT-PCR, 2) determine genes correlated with *CTGF* mRNA levels in three independent pre-B ALL cohorts, and compare their promoters with respect to common regulatory motifs, and 3) examine

whether *CTGF* expression is linked to diagnostic features or prognosis in paediatric pre-B ALL.

3.2 Results

3.2.1 The PMH cohort of paediatric pre-B ALL

The global gene expression profiles of primary pre-B ALL specimens was measured by Boag and colleagues using Affymetrix microarray technology (Boag et al. 2006; Boag et al. 2007). This group of specimens, named the PMH cohort, comprised 73 paediatric pre-B ALL specimens from patients treated on COG protocols (Gaynon et al. 2000). These were obtained either at the time of diagnosis; n=60, or at relapse; n=13. The critical clinical and biological features of this cohort are detailed in Table 3.1.

Prior to analysing the global gene expression data from the PMH cohort (Boag et al. 2006; Boag et al. 2007), *CTGF* expression levels measured by microarray were validated using 17 primary pre-B ALL specimens from this cohort. Also included were 3 T-ALL specimens for comparison. The details of these specimens are outlined in Table 3.2. These specimens, also used in later parts of the current study (Chapters 4 & 6) were assayed for *CTGF* expression by quantitative real-time PCR (qRT-PCR). RNA from these 20 primary bone marrow samples was extracted from cryopreserved material that was matched with those used for the previous microarray study (Boag et al. 2007). These specimens represented a range of cytogenetic subtypes, determined using a previously published 20 gene classifier model for cytogenetic prediction (Hoffmann et al. 2006). These were *MLL* (n=1), *BCR-ABL* (n=4), hyperdiploidy (n=2), *TEL-AML* (n=2), *TCF3-PBX1* (n=1) and 10 cases where known cytogenetic abnormalities could not be detected (Other). These latter specimens were all classed as low/neg for *CTGF* mRNA by microarray as they exhibited RMA normalised probeset signals of <50 (see section 2.2.1.1).

Table 3.1 Clinical and biological features of pre-B ALL patients in the PMH cohort

	Number	%		Number	%
Age			CNS involvement		
1-10 years	53	73	Not known	4	5.5
>10 years	20	27	Present	1	1.5
Gender			Absent	68	93
Male	41	56	Spleen		
Female	32	44	Normal	32	44
WBC[§]			Enlarged (BU) ‡	37	51
<50 x 10 ⁹ /l	67	92	Enlarged (AU) ‡	4	5
>50 x 10 ⁹ /l	6	8	Lymph nodes		
NCI risk *			Normal	37	51
Standard	50	68	Enlarged <3cm	32	44
High	23	32	Enlarged >3cm	4	5
Specimens			Mediastinal mass		
Diagnosis	60	82	Not known	12	16
Relapse	13	18	No mass	58	80
Clinical outcome			Anterior mass	3	4
(diagnosis specimens)	(n=60)				
CCR †	11	18			
Relapsed	49	82			
Cytogenetics #					
<i>MLL</i>	1	1			
<i>BCR-ABL1</i>	7	10			
<i>TEL-AML1</i>	15	21			
<i>TCF3-PBX1</i>	2	3			
Hyperdiploid	26	36			
Other	22	30			

§ WBC: Peripheral white blood cell count at diagnosis, † CCR: Complete clinical remission, * NCI risk stratification: see Smith et al., (1996), # Cytogenetics: determined by karyotype or inferred by gene expression profiling, see Hoffman et al., (2006), ‡ BU/AU: below/above umbilicus.

Table 3.2 Characteristics of pre-B ALL specimens used for validation of *CTGF* expression

The panel of primary specimens used in this study consisted of 17 pre-B ALL, both diagnosis (D) and relapse (R) specimens, and 3 T-ALL specimens. *CTGF* expression measured by Affymetrix (U133A) human genome array and classed as high, med or low/negative based on probeset intensity (defined in Section 2.2.1.1).

Specimen Code	Diagnosis /Relapse	Lineage	cytogenetics (inferred) #	NCI Risk *	Age ‡ (months)	<i>CTGF</i> expression
BHDN9	D	B	<i>MLL</i>	High	166	high
[±] BHDR25	D	B	<i>BCR-ABL</i>	High	196	high
[±] BHRR25	R	B	<i>BCR-ABL</i>	High	201	high
BHDN12	D	B	HYP	High	29	high
BSDR85	D	B	<i>TEL-AML</i>	Standard	92	high
BSDN36	D	B	<i>TEL-AML</i>	Standard	64	high
BSDR79	D	B	<i>BCR-ABL</i>	Standard	113	med
BHRR83	R	B	HYP	High	250	med
BSDN34	D	B	<i>BCR-ABL</i>	Standard	50	med
BSRR89	R	B	<i>TCF3-PBX1</i>	Standard	133	med
BSDN67	D	B	Other	Standard	79	low/neg
BSDN40	D	B	Other	Standard	97	low/neg
BHDN7	D	B	Other	High	132	low/neg
BSDN51	D	B	Other	Standard	74	low/neg
BHDR31	D	B	Other	High	171	low/neg
BHRR92	R	B	Other	High	177	low/neg
BSDN63	D	B	Other	Standard	30	low/neg
THDN2	D	T	Other	High	70	low/neg
THDN4	D	T	Other	High	59	low/neg
THDN6	D	T	Other	High	161	low/neg

[±] Paired diagnosis and relapse specimens, # Cytogenetics inferred by 20 gene classifier method (Hoffmann et al. 2006), * NCI risk stratification, see Smith et al., (1996), ‡ Age in months at the time of diagnosis or relapse.

3.2.2 Validation of *CTGF* expression data

CTGF expression values generated by microarray (probeset 209101_at) and qRT-PCR were examined for concordance in these 20 ALL specimens. Figure 3.1 demonstrates *CTGF* gene expression levels and the correlation between the methods used to characterise expression in these samples. A comparison of *CTGF* expression values generated by microarray versus those obtained by qRT-PCR performed on freshly extracted, matched cryopreserved specimens, demonstrated a highly similar pattern of *CTGF* expression. Specimens BSDN36 and BSDR79 seemed to display different *CTGF* expression levels when measured by qRT-PCR, however overall a Pearson correlation of 0.972 was observed between methods, which was highly significant ($p < 0.001$). Those primary specimens that lacked one of the common cytogenetic rearrangements (Table 3.2 “Other”), were considered to be *CTGF*^{low/neg} when assessed by microarray, however, almost all of these cells exhibited an extremely low level of *CTGF* mRNA detectable by qRT-PCR, indicating the *CTGF* locus is in a permissive state.

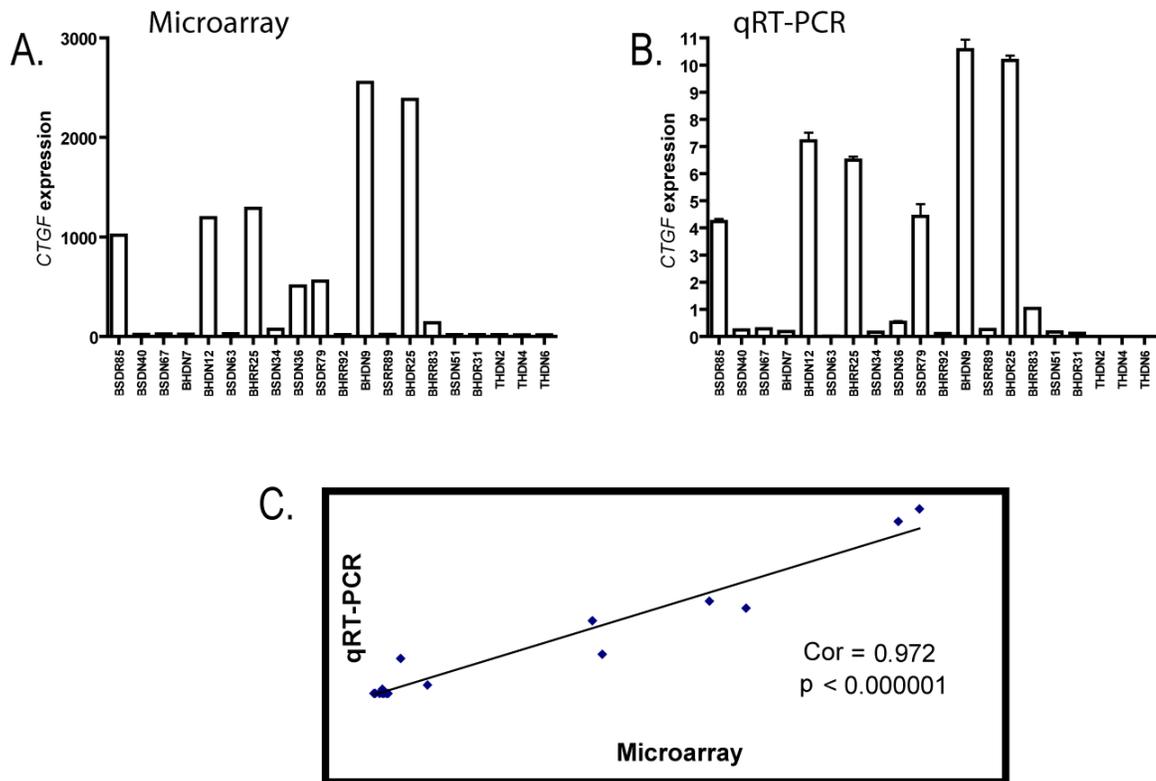


Figure 3.1 Analysis of *CTGF* expression by microarray and qRT-PCR

A. *CTGF* expression in a selection of primary B-lineage (n=17) and T-lineage (n=3) ALL specimens measured by Affymetrix human genome array U133A (probe 209101_at) **B.** *CTGF* expression measured in matched, cryopreserved specimens by qRT-PCR normalised to *ACTB*. Error bars represent the standard error of the mean of three technical replicates. **C.** Comparison of *CTGF* expression data generated by microarray and qRT-PCR. A highly significant Pearson correlation of 0.972 was observed (p<0.001) between the two methods.

3.2.3 Gene expression profiles associated with *CTGF* expression

Identifying genes that are co-regulated with *CTGF* may point to common promoter elements or trans-acting factors involved in activation of the *CTGF* promoter (Mansson et al. 2004). Probesets that correlated with *CTGF* expression in pre-B ALL were identified by examining global gene expression data from the PMH cohort (n=73), and two publically available paediatric pre-B ALL cohorts. These two cohorts were denoted

as the Ross cohort and the Kang cohort (Ross et al. 2003; Kang et al. 2010). The former was studied by Ross and colleagues while investigating the suitability of gene expression profiling for classifying paediatric ALL specimens into clinically relevant subtypes (Ross et al. 2003). This cohort consisted of 118 pre-B ALL specimens, *BCR-ABL1*; n=15, *TCF3-PBX1*; n=18, Hyperdiploid; n=15, *MLL* disease; n=20, *TEL-AML1*; n=20, and 28 cases where the aforementioned cytogenetic subtypes were not detected (Ross et al. 2003). The Kang cohort was comprised of 207 high-risk paediatric pre-B ALL patients enrolled in the Children's Oncology Group (COG) P9906 clinical trial and defined by age > 10yrs, WBC > $5 \times 10^4/\mu\text{l}$, and the absence of low-risk (Hyperdiploidy, *ETV6-RUNX1*) or very high-risk (hypodiploid, *BCR-ABL1*) disease (Kang et al. 2010). However, patients with low-risk cytogenetics were included if they displayed high-risk features like CNS or testicular involvement. While the Kang cohort was comprised of a discreet, high-risk group, it was considered appropriate for investigating *CTGF* correlates because *CTGF* was shown to be highly predictive of outcome in their study (Kang et al. 2010).

The distribution of *CTGF* gene expression values was assessed across samples in each cohort without normalizing the three datasets together, as the Kang cohort had been analysed using a different Affymetrix microarray platform (U133 plus2.0 array) to the PMH and Ross cohorts (U133A array). The range of *CTGF* expression in each cohort is shown in Figure 3.2A. A Spearman correlation analysis was performed on each cohort individually to examine which probesets displayed a similar pattern of expression to *CTGF*. The top 100 ranked probesets from the Spearman correlations were compared among cohorts and 11 probe sets overlapped between cohorts, as illustrated in Figure 3.2B. These probesets corresponded to 6 genes; suppressor of cytokine signalling 2 (*SOCS2*), myocyte enhancer factor 2C (*MEF2C*), adducin 3 (*ADD3*), gelsolin (*GSN*), metastasis suppressor 1 (*MTSS1*) and dihydropyrimidinase-like 2 (*DPYSL2*). These probesets displayed significant Spearman correlation values in the PMH cohort ranging from 0.68 ($p < 10^{-7}$) to 0.51 ($p < 10^{-5}$) and are shown in Figure 3.2C. The variance in the expression of these 11 probesets was compared to confirm that they display a similar pattern as *CTGF* (Figure 3.2D). All but one of the probesets, corresponding to *SOCS2*,

MEF2C, *ADD3*, *GSN* and *DPYSL2* appeared to have a distribution very similar to *CTGF* suggesting they may be co-regulated. In contrast, *MTSS1* gene expression varied substantially from *CTGF* in that it was expressed at comparatively lower levels and displayed less variance in expression among samples. As such this gene was omitted from further analysis.

Based upon the significant correlation with *CTGF* expression displayed by *SOCS2*, *MEF2C*, *ADD3*, *GSN* and *DPYSL2*, it was considered that there may be common *cis* regulatory elements in the 5' flanking region of these genes. One of these co-regulated genes, *MEF2C*, encodes a transcription factor, suggesting that it may itself be a candidate. However, analysis of the promoter regions of the above genes using the TFSEARCH database of transcription factor motifs available at <http://www.cbrc.jp/research/db/TFSEARCH.html>, did not identify any predicted *MEF2C* binding sites, even when the region examined was extended to include up to 4kb of 5' flanking sequence adjacent to the transcription start site.

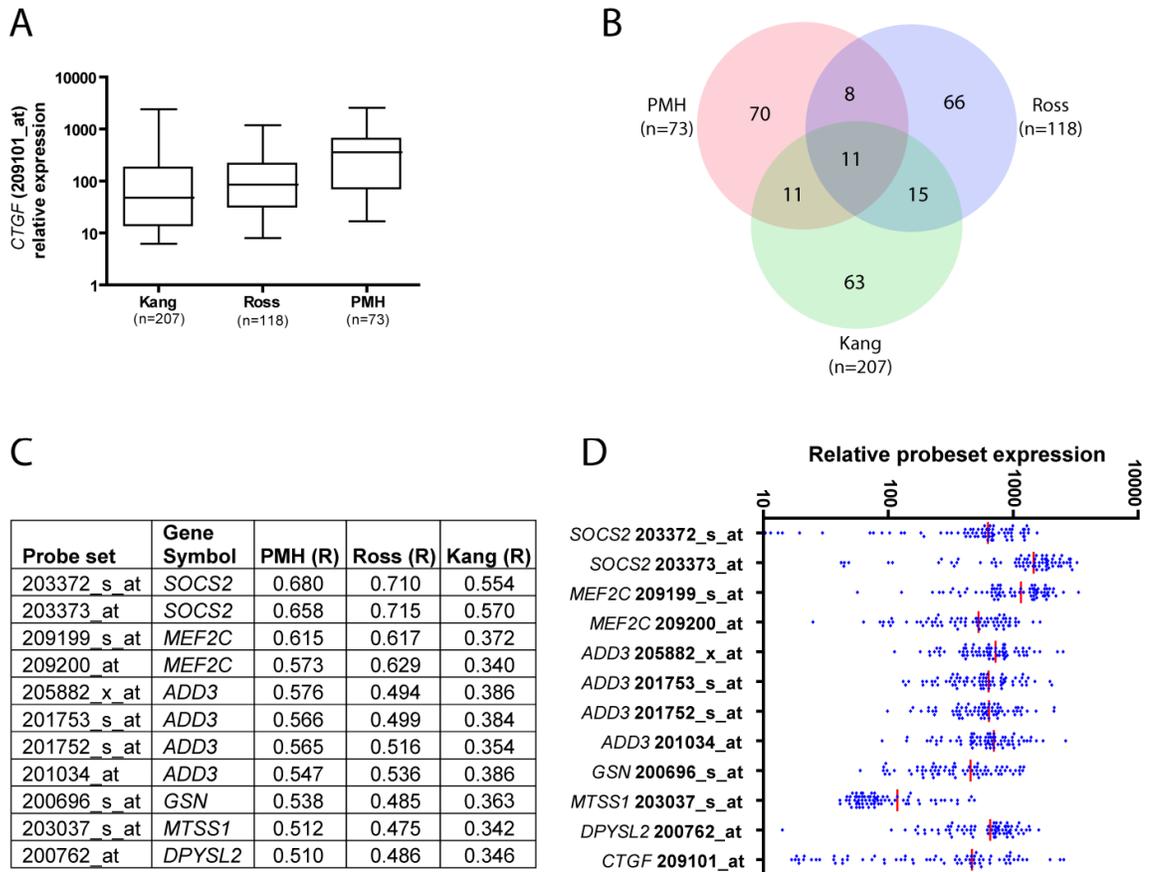


Figure 3.2 CTGF correlation analysis in three independent cohorts

A. Distribution of *CTGF* expression (probeset 209101_at) measured by microarray in three independent paediatric pre-B ALL cohorts. Cohort datasets are not normalised together as they originate from different Affymetrix platforms. **B.** A Spearman correlation analysis with *CTGF* (209101_at) was performed in each cohort individually and the top 100 probesets from each cohort were compared. Eleven common probesets were identified in the top 100 probesets in each cohort. **C.** Overlapping probe sets corresponded to 6 genes; suppressor of cytokine signalling 2 (*SOCS2*), myocyte enhancer factor 2C (*MEF2C*), adducin 3 (*ADD3*), gelsolin (*GSN*), metastasis suppressor 1 (*MTSS1*) and dihydropyrimidinase-like 2 (*DPYSL2*). For each probeset the correlation with *CTGF* expression (R) is shown for each cohort. **D.** Expression of the 11 overlapping top-ranked probesets as well as *CTGF* in the PMH cohort measured by microarray.

3.2.2.1 Common 5' flanking motifs in *CTGF*-correlated genes

In order to identify common sequence elements, the 5' flanking sequence of *CTGF* together with five highly correlated genes from -2 kb to +1 (relative to the transcriptional start site) were examined for shared sequence motifs using the MEME software suite (Bailey & Elkan 1994) available at <http://meme.nbcr.net/meme/intro.html>. This program searches for common sequence motifs in specified DNA sequences. Analysis with MEME identified two sequence motifs that were present in the 5' flanking sequence of the majority of the genes determined to be co-regulated with *CTGF*. These two motifs were referred to as MEME enriched motif 1 (MEM1) and MEM2.

The first motif, MEM1, was 15bp in length, heavily G-rich and was present in the promoters of *ADD3*, *DPYSL2*, *GSN*, *CTGF* and *SOCS2* (Figure 3.3A). This motif was examined for transcription factor binding sites using the TOMTOM motif comparison tool (Gupta et al. 2007) available as part of the MEME suite of applications. This program compares sequence motifs against known transcription factor binding matrices annotated in the TRANSFAC (Fu & Weng 2004) and JASPER databases (Portales-Casamar et al. 2010). The results of this analysis is shown in Figure 3.3B and predicted a binding site for the Ikaros family of transcription factors in MEM1 which are known to play an important role in haemopoietic development (Koipally et al. 1999). Although the Ikaros-binding site was predicted by the TOMTOM analysis, the Ikaros-binding motif was only a partial match to the 5' region of MEM1. Furthermore, the location of the MEM1 motif in the 5' flanking sequences shown in Figure 3.3 was highly variable, suggesting that this motif may not be a functional element, but rather a simple sequence repeat identified by the MEME analysis.

The second motif, MEM2, was 32 bp in length, as shown in Figure 3.4 and was identified at similar positions (approximately -650 to -1150) in the promoter regions of *ADD3*, *SOCS2*, *GSN* and *DPYSL2*, but was not present in the *CTGF* promoter. Significantly, this motif was also predicted to facilitate binding of the Ikaros family of

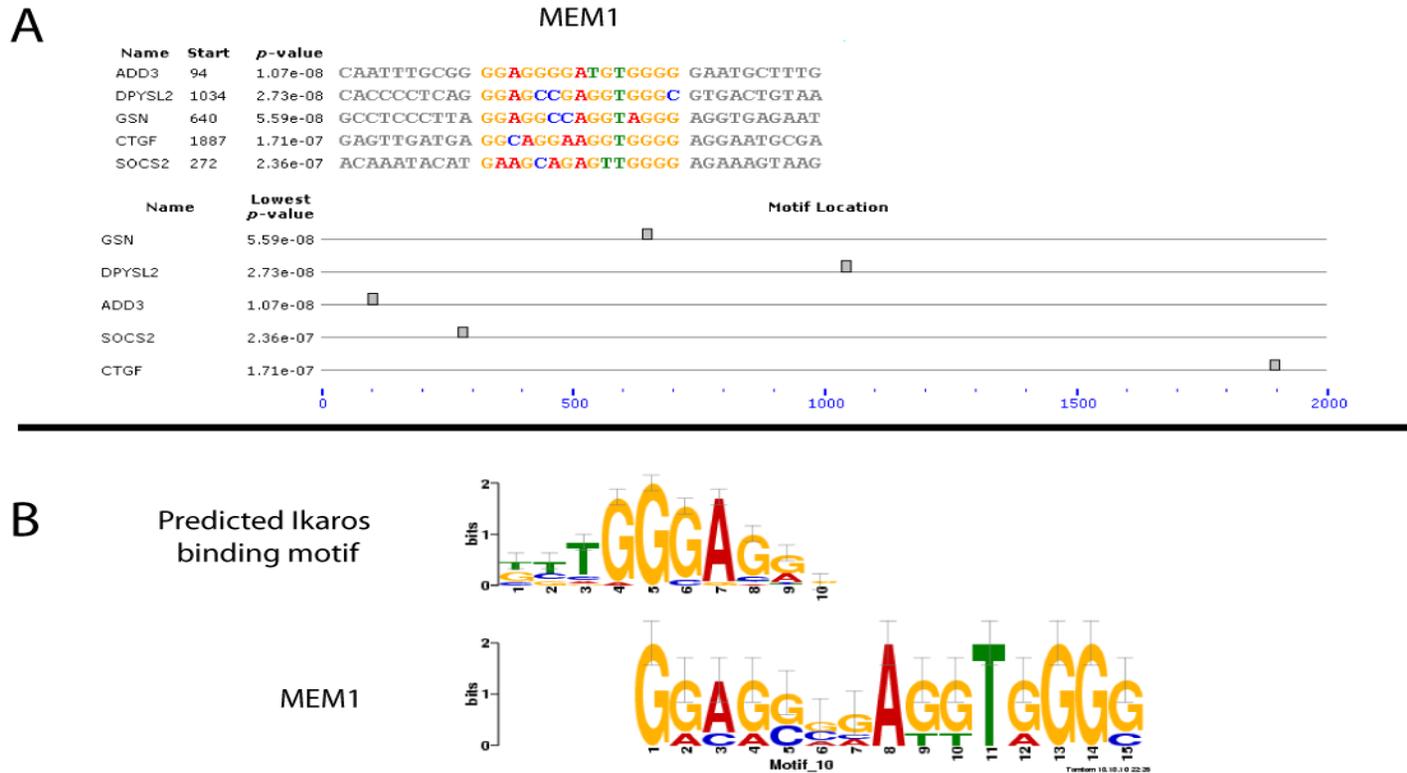


Figure 3.3 MEME enriched motif 1 (MEM1)

A. Top; Meme enriched motif (MEM) 1 and flanking sequence for each promoter including gene symbol (name), nucleotide coordinate (start) of motif in input sequence (-2kb to +1) and p-value of motif discovery. Sequences are ranked from top to bottom by statistical significance. Bottom; location of motif in input sequence. **B.** Top; Predicted Ikaros-binding motif. Bottom; MEM1 and ideogram representing nucleotide incidence at each base in motif. Analysis was performed using MEME and TOMTOM programs available online at <http://meme.nbcr.net/meme/intro.html>.

A

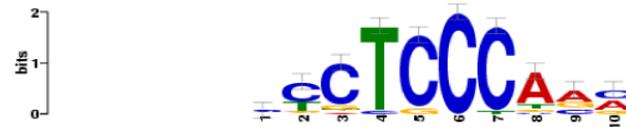
MEM2

Name	Start	p-value	Sequence
ADD3	1341	8.25e-30	GATCTACCTG CCTCAGCCTCCCAAAGTGCTGGGATTACAGGCCTGAGCCACCACGCCCGG CAATATTGTG
SOCS2	843	3.40e-28	GATCCACCCG CCTCAACCTCCCAAAGTGCTGGGATTACAGGCCTGAGCCACCACGCCCGG CCAGGATTCT
GSN	1060	5.03e-20	CCGGAGTCAC CCGGCAACTCCAAAAGTCTGGGAACTCATGTGTGCACCAGGACTGCTGG GAGGCCGTCT
DPYSL2	941	4.04e-19	CACTGTAAC CCTCAGGACCCCGGGCGCTGGGATCGCAGAGCTGCCCCGACAGCGCT GCAGGCACCA



B

Predicted Ikaros binding motif



MEM2



Figure 3.4 MEME enriched motif 2 (MEM2)

A. Top; Meme enriched motif (MEM) 2 and flanking sequence for each promoter including gene symbol (name), nucleotide coordinate (start) of motif in input sequence (-2kb to +1) and p-value of motif discovery. Sequences are ranked from top to bottom by statistical significance. Bottom; location of motif in input sequence. **B.** Top; Predicted Ikaros-binding motif. Bottom; MEM2 and ideogram representing nucleotide incidence at each base in motif. Analysis was performed using MEME and TOMTOM programs available online at <http://meme.nbcr.net/meme/intro.html>.

transcription factors and displayed higher homology with the IKAROS motif than MEM1. This motif additionally contained an invariant CTGGGA sequence in all four promoters, however this sequence element was not predicted to be part of any transcription factor binding motif by TOMTOM. This was surprising, as the first three members of the Ikaros family; IKAROS, HELIOS and AIOLOS, all contain a central GGGGA sequence in their DNA recognition sequences (shown in Figure 3.5) and the invariant CTGGGA identified in MEM2 is compatible with IKAROS, HELIOS and AIOLOS binding matrices (Molnar & Georgopoulos 1994). While the data provided by analysis with MEME and TOMTOM were novel, the existence of Ikaros family binding sites in these genes warranted further validation.

IKZF1 - IKAROS						IKZF2 - HELIOS						IKZF3 - AIOLOS					
pos	A	C	G	T	consensus	pos	A	C	G	T	consensus	pos	A	C	G	T	consensus
1	7	5	7	5	N	1	9	10	12	5	N	1	3	0	7	15	T
2	6	9	0	9	N	2	11	5	7	13	N	2	8	8	5	4	N
3	3	3	7	11	N	3	5	7	7	17	N	3	0	13	3	9	Y
4	2	6	1	15	T	4	6	10	0	20	Y	4	6	5	0	14	T
5	2	0	22	0	G	5	6	0	30	0	G	5	5	0	20	0	G
6	0	0	24	0	G	6	0	0	36	0	G	6	0	0	25	0	G
7	0	0	24	0	G	7	0	0	36	0	G	7	0	0	25	0	G
8	24	0	0	0	A	8	36	0	0	0	A	8	25	0	0	0	A
9	22	0	0	2	A	9	18	5	1	12	W	9	25	0	0	0	A
10	3	4	2	15	T	10	10	7	6	13	N	10	0	4	0	18	T
11	11	2	8	3	R	11	12	8	9	7	N	11	16	0	5	4	A
12	3	13	2	6	C	12	9	18	5	4	N	12	1	18	0	6	C
13	3	14	3	4	C							13	1	17	0	7	C

Figure 3.5 DNA binding matrices for the Ikaros family members 1, 2 and 3

DNA binding matrices are shown for the first three members of the Ikaros family of transcription factors. These members are IKAROS (IKZF1), HELIOS (IKZF2) and AIOLOS (IKZF3). Figure adapted from Molnar et al., (1994).

3.2.2.2 Analysis of putative Ikaros family binding sites

The *In silico* evidence presented above suggested that members of the Ikaros family of transcription factors IKAROS, HELIOS and AIOLOS may bind to the 5' flanking sequence of genes displaying similar patterns of expression as *CTGF*. Each promoter was independently examined for predicted Ikaros family binding sites using the TFSEARCH transcription factor motif database (<http://www.cbrc.jp/research/db/TFSEARCH.html>). The 5' flanking sequence of *GSN*,

DPYSL2, *ADD3*, *SOCS2*, *MEF2C* and *CTGF* were all predicted to contain at least one, and in some cases several, binding sites for the Ikaros family member HELIOS as outlined in Figure 3.6. A number of sites co-localised with the matrices predicted by MEME in Figures 3.3 and 3.4. These sites, annotated in Figure 3.6 included 1, 2, 3, 4, 8, 10 and 12. These predicted binding sites are based on functionally characterised murine Helios binding motifs (Molnar & Georgopoulos 1994), however the sequence identity shared between the human and murine *HELIOS/Helios* coding region is over 97% and they are predicted to recognise similar, if not identical, recognition sequences (Hosokawa et al. 1999). Furthermore, some of the predicted binding sites were also compatible with binding by IKAROS. The remarkable similarities in the binding matrices for IKAROS, HELIOS and AIOLOS as outlined in Figure 3.5 suggested that while some predicted binding sites are only compatible with HELIOS or IKAROS, any of these three proteins may in fact bind to the sites identified in the promoters of genes that showed expression correlated to *CTGF* gene expression.

3.2.2.3 Expression of Ikaros family genes in the PMH cohort

Binding motifs for HELIOS and IKAROS were identified in the 5' flanking sequence of genes that correlate with *CTGF* expression, suggesting a role for the Ikaros family of transcription factors in regulating expression of these genes. The PMH microarray dataset was examined to determine whether changes in expression of any Ikaros family members; *IKAROS*, *HELIOS*, *AIOLOS*, *EOS* and *PEGASUS* was associated with *CTGF* expression in this cohort (Figure 3.7). Specimens were grouped into *CTGF*^{pos} and *CTGF*^{low/neg} groups based on microarray expression values, and mean expression of the Ikaros family of genes were compared between these groups. All Ikaros family genes were expressed except for *AIOLOS* which displayed probeset intensity values well below what would normally be classed as a background signal on the U133A chip (intensity value of <25). There was no significant difference observed in the mean expression of any Ikaros family member between *CTGF*^{pos} and *CTGF*^{low/neg} specimens as shown in Figure 3.7, suggesting that any effect mediated by HELIOS or other Ikaros proteins was not due to alterations in their mRNA expression levels.

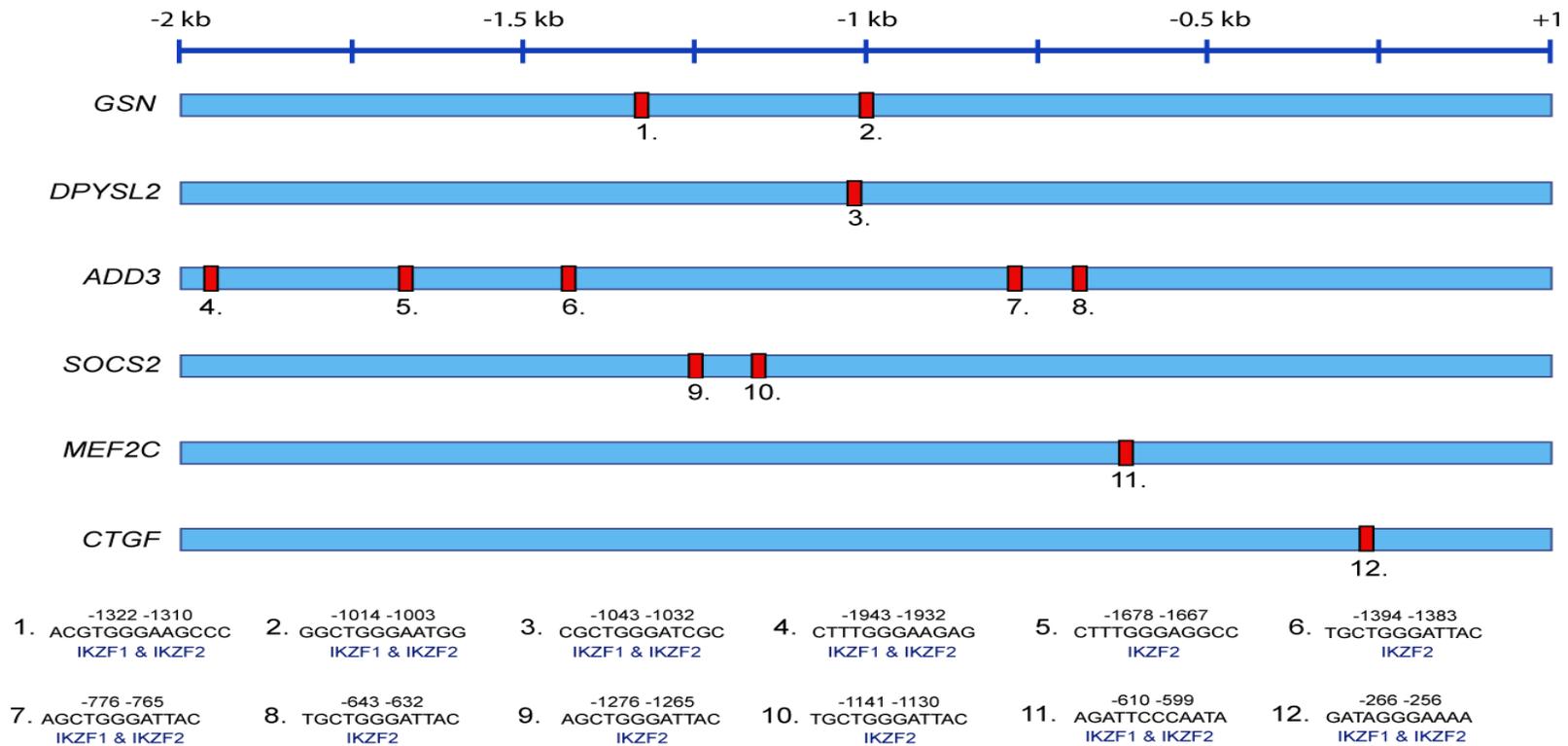


Figure 3.6 Helios (IKZF2) is predicted to bind the promoters of *CTGF* and correlated genes

Binding sites for IKZF1/IKAROS and IKZF2/HELIOS predicted in the 5' flanking regions (-2 kb to +1) of *CTGF*-correlated genes by TFSEARCH. Numbered red boxes indicate predicted binding sites (1-12) which are annotated (bottom) with coordinates relative to the transcription start site of each gene, motif sequence, and indication of which factor was predicted to bind the target.

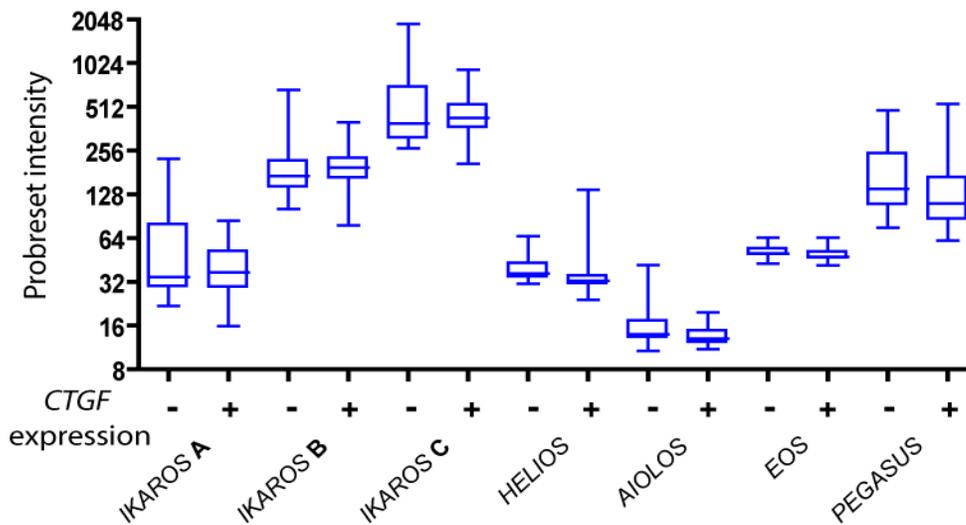


Figure 3.7 Expression of Ikaros family members in the PMH cohort

Expression of Ikaros transcription factors, measured by microarray analysis of primary specimens in the PMH cohort. Specimens were grouped according to *CTGF* expression status (samples classed as *CTGF*^{low/neg} are denoted by the minus symbol). Expression indicated by probeset intensity for *IKAROS* probesets; A-216901_s_at, B-205038_at and C-205039_s_at, as well as *HELIOS* 220567_at, *AIOLOS* 221092_at, *EOS* 208472_at and *PEGASUS* 220086_at.

3.2.2.4 *IKAROS* gene lesions are associated with *CTGF* expression

Deletions and mutations of *IKAROS* are common events in high-risk pre-B ALL (Kuiper et al. 2010; Mullighan et al. 2008), and internal in-frame deletions can produce dominant-negative (DN) protein isoforms that can inhibit the function of Ikaros proteins, in the absence of changes in gene expression (Sun et al. 1996). The Ikaros family of proteins is thought to promote gene silencing through enforcing repressive chromatin marks at target gene loci (Koipally et al. 1999). If aberrant *CTGF* expression can indeed be attributed to a loss of function in Ikaros family genes in pre-B ALL, then those specimens harbouring mutations or deletion of *IKAROS* would be expected to exhibit high *CTGF* expression. A recent report from St Jude Children’s Research Hospital

examined the prognostic significance of *IKAROS* deletions and mutations, herein referred to collectively as *IKAROS* alterations, using a combination of SNP arrays and direct sequencing of B-cell developmental genes (Mullighan et al. 2009). This study was undertaken in a selection of high-risk patients from the COG P9906 clinical trial cohort. Patient specimens used for the study by Kang and colleagues were also selected from the COG P9906 cohort (Kang et al. 2010). Despite the selection criteria of the two studies being very similar, focussing on high-risk pre-B ALL, not all specimens examined for copy number alterations by Mullighan *et al* were represented in the gene expression data associated with the Kang cohort. However, of the 221 specimens in the Mullighan cohort, 198 matched with gene expression data from the Kang cohort. Pairing of this data allowed the exploration of the effect of *IKAROS* alterations upon *CTGF* expression in high-risk pre-B ALL.

Of the 67 patient specimens that were identified as harbouring an *IKAROS* alteration by Mullighan *et al*, gene expression data from the Kang cohort was available for 59 of these. To examine whether *IKAROS* alterations were associated with aberrant *CTGF* expression, mean *CTGF* expression was compared between these 59 specimens and the remainder of specimens in the Mullighan cohort for which gene expression data was available. This analysis shown in Figure 3.8A demonstrated that *IKAROS* alterations were indeed associated with higher *CTGF* expression, with a significant difference in mean *CTGF* expression evident between these two groups ($p=0.036$).

CTGF expression was also investigated in a more homogeneous subset of pre-B ALLs, which all contained a deletion in at least one critical B-cell developmental gene specimens (66.5 % of high-risk cases), including *IKAROS*, *PAX5*, *TCF3*, *EBF1*, *RAG1/2*, *BLNK*, *BCL11A*, *HELIOS*, *LEF1*, *MEF2C*, *SOX4* and *SPI1* (Mullighan et al. 2009). *IKAROS* deletions were present in 43% of this group compared to 29% in the total P9906 cohort, thus it was expected that any association between *IKAROS* alterations and *CTGF* gene expression would be easily observed in this group. These specimens were stratified based upon the presence or absence of *IKAROS* alterations and examined for *CTGF* expression, shown in Figure 3.8B. A total of 154 specimens

harboured disruption of one or more of these critical B-cell developmental genes. Of these 154 specimens, gene expression data was available for 137 specimens, and 59 of these were affected by *IKAROS* alterations. This group of 59 specimens exhibited higher mean *CTGF* expression compared to those specimens harbouring B-cell pathway lesions but no *IKAROS* alterations and this was highly significant ($p < 0.001$). These findings support the hypothesis that *IKAROS* alterations are associated with aberrant *CTGF* expression in high-risk pre-B ALL.

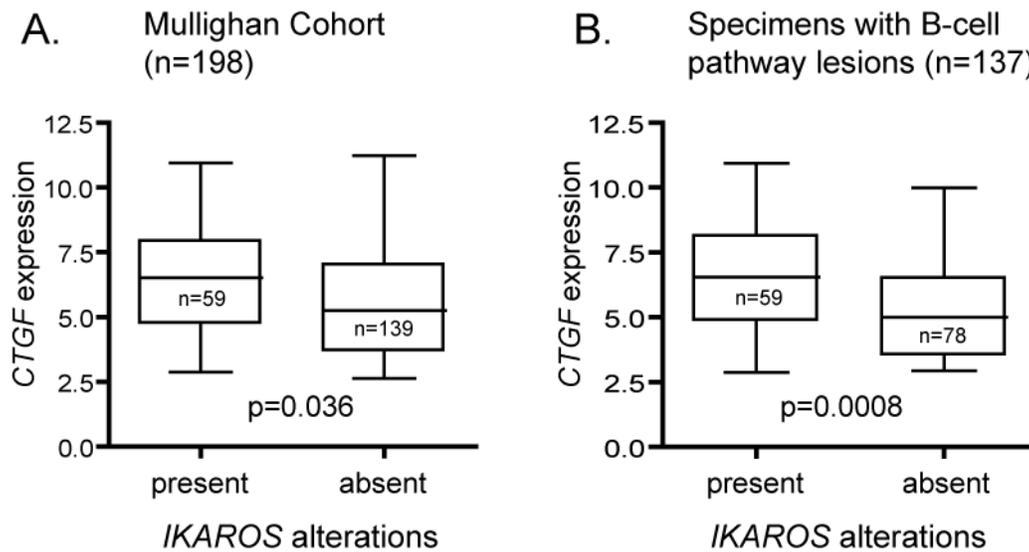


Figure 3.8 *IKAROS* alterations are associated with aberrant *CTGF* expression in high-risk pre-B ALL

A. Patient specimens from the Mullighan cohort (Mullighan et al. 2009), were divided into two groups based on whether they had a confirmed *IKAROS* alteration, then examined for expression of *CTGF*. Those patients with alterations affecting *IKAROS* (n=59) displayed significantly higher *CTGF* expression ($p=0.036$) compared to the remainder of the cohort in which alterations of the *IKAROS* locus were absent. **B.** *CTGF* expression was examined in patient specimens harbouring disruption of at least one critical B-cell pathway gene, including *IKAROS*, *PAX5*, *TCF3*, *EBF1*, *RAG1/2*, *BLNK*, *BCL11A*, *HELIOS*, *LEF1*, *MEF2C*, *SOX4* and *SPI1*. Of the 137 specimens harbouring mutation of at least one of these genes, 59 had a deletion or mutation affecting the

IKAROS gene, and these specimens had significantly higher *CTGF* expression ($p < 0.001$).

3.2.3 Clinical features associated with *CTGF* expression

CTGF expression reportedly has prognostic significance in pre-B ALL and has been associated with reduced RFS for some forms of the disease in both adults and children (Sala-Torra et al. 2007; Kang et al. 2010). As shown in Figure 3.9, *CTGF* expression in the PMH cohort varied considerably across specimens. This suggested that *CTGF* expression might correlate with one or more clinical features present at the time of diagnosis, thus acting as a biomarker for a specific group (defined by age, cytogenetic group, etc) at greater risk (Yoshida et al. 2009; Choi et al. 2006). Patients comprising the PMH cohort were diagnosed locally, enabling access to detailed patient data for this study.

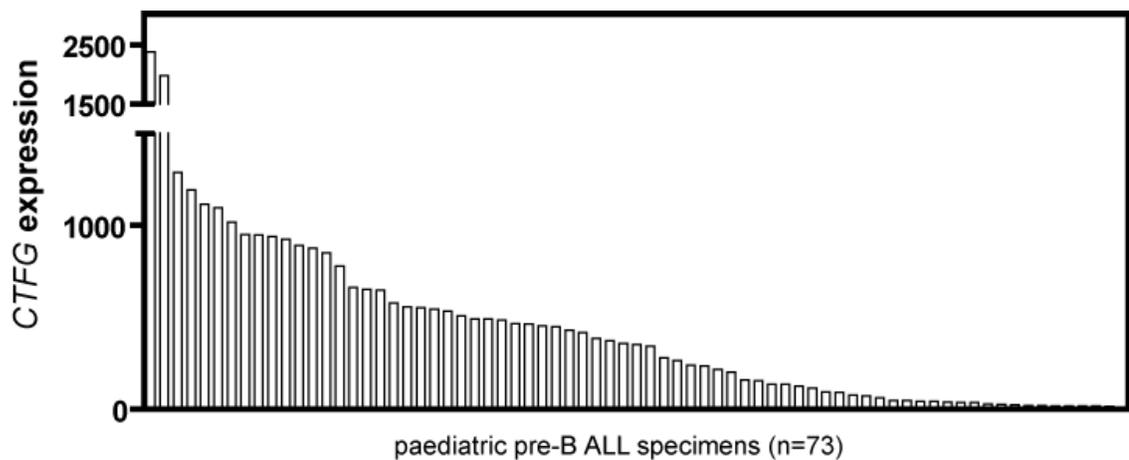


Figure 3.9 *CTGF* gene expression in the PMH cohort

CTGF gene expression measured by Affymetrix human genome array U133A (probeset 209101_at) in the paediatric pre-B ALL cohort from PMH (n=73).

Patient specimens obtained at the time of diagnosis (n=60) were stratified based upon clinical features including gender (male; n=33, female; n=27), age at diagnosis in years (1.5-3 yrs; n=12, 3-6 yrs; n=16, 6-10 yrs; n=22 and 10-20 yrs; n=10), percentage blasts in bone marrow (>95%; n=38, 90-95%; n=12 and <90%; n=10), haemoglobin levels in g/dL, (>10; n=12, 7.5-10; n=18, 7.5-5; n=25 and <5; n=5), lymph node status (enlarged; n=33, normal; n=27) and spleen status (enlarged; n=38, normal; n=22). *CTGF* expression was also compared between these 60 diagnosis specimens and 13 relapse specimens. The mean *CTGF* expression was compared between these clinically relevant groups as shown in Figure 3.10. Patients with enlarged lymph nodes displayed significantly lower *CTGF* expression (p=0.0349) than those that did not have lymph node involvement. A similar trend was observed for those patients with an enlarged spleen, however this did not reach significance. No other clinical features examined demonstrated an association with *CTGF* expression.

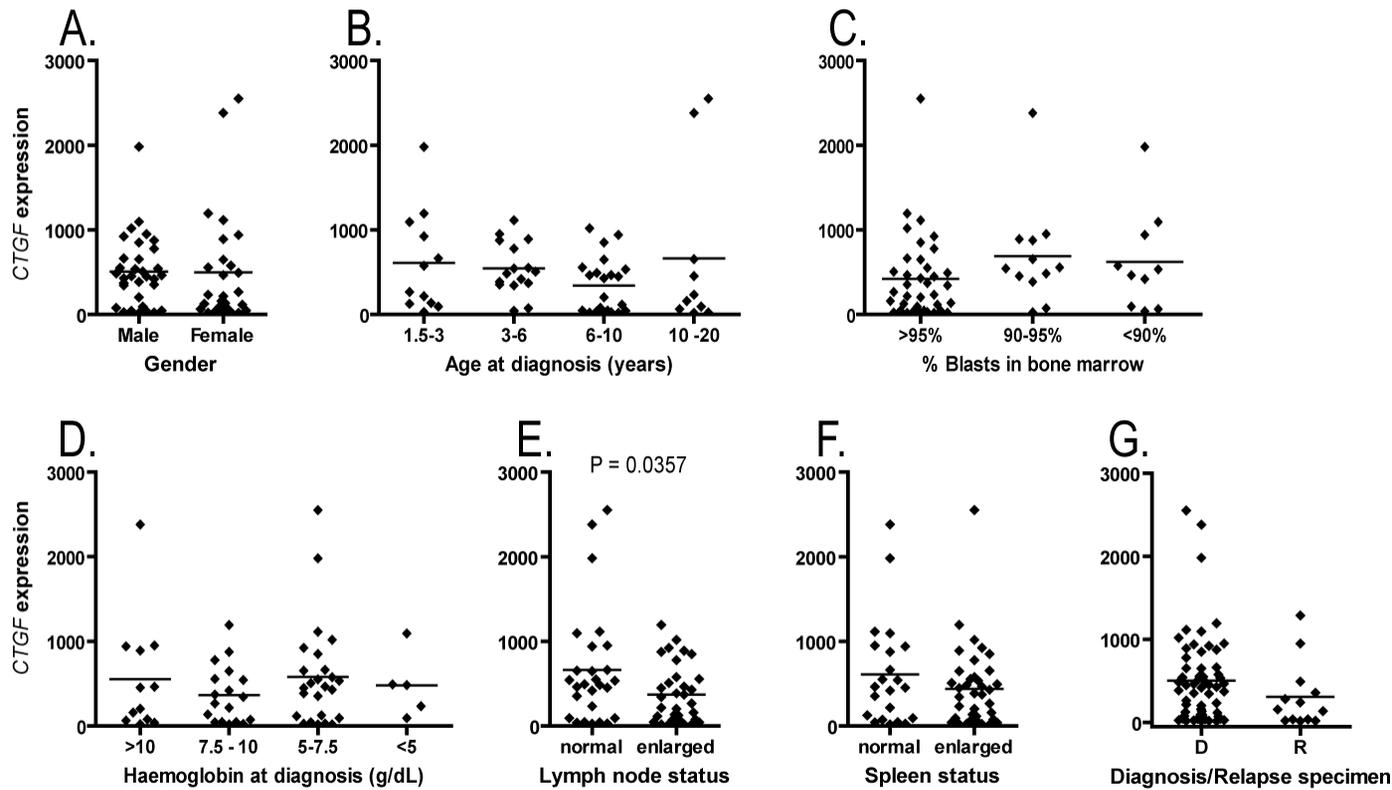


Figure 3.10 Comparison of *CTGF* expression versus clinical features present at diagnosis in the PMH cohort

CTGF gene expression (measured by microarray) in pre-B ALL specimens (n=60) stratified by clinical features present in patients at the time of diagnosis, including gender (A.), age at diagnosis (B.), percentage blasts in bone marrow (C.), haemoglobin (D.), lymph node status (E.) and Spleen status (F.). *CTGF* expression was also compared between the 60 diagnosis specimens and 13 relapse specimens (G.). Mean *CTGF* expression was compared between groups using unpaired t-tests or one-way ANOVA where appropriate.

3.2.4 Prognostic significance of *CTGF* expression

The association between *CTGF* expression in specimens obtained at diagnosis in the PMH cohort and both patient relapse and overall survival was examined to investigate the prognostic relevance of *CTGF* expression (Figure 3.11). Patient diagnostic specimens were stratified based upon *CTGF* expression into two groups, *CTGF*^{pos} and *CTGF*^{low/neg}, based on previously determined gene expression cut-offs (section 2.2.1.1), and the incidence of relapse and overall survival was examined. Three patients were censored from the analysis of overall survival due to incomplete survival data for the 5yr period. There was a trend toward lower RFS in the *CTGF*^{pos} group at 71% compared to 83% for the *CTGF*^{low/neg} group, however this was not statistically significant (p=0.39). The overall 5 year survival did not differ between the *CTGF*^{pos} and *CTGF*^{low/neg} groups. The trend for higher relapse lends support to previous studies indicating that *CTGF* has prognostic significance in pre-B ALL (Sala-Torra et al. 2007; Kang et al. 2010).

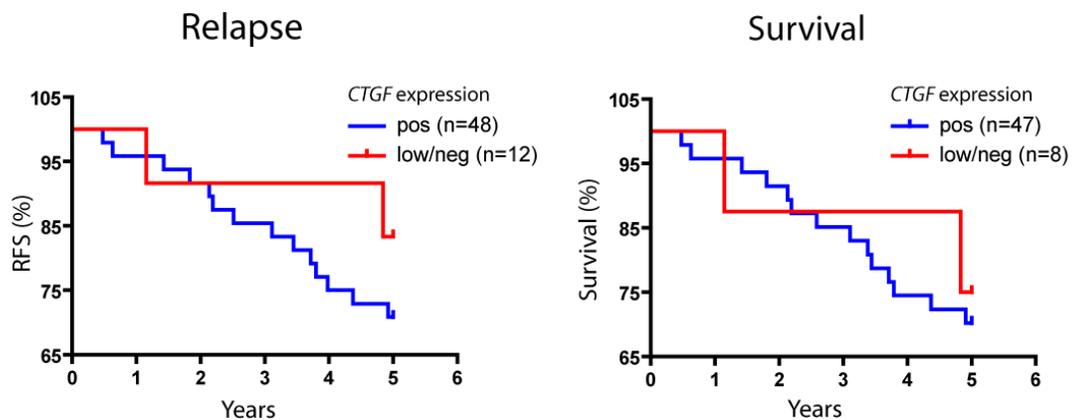


Figure 3.11 Association between *CTGF* expression and patient outcome.

Patients were stratified based upon *CTGF* expression (pos or low/neg) and assessed for both relapse-free survival (RFS) and overall survival over 5 years.

3.3 Discussion

Deregulated *CTGF* expression in pre-B ALL has now been identified by several studies, however the mechanisms driving its aberrant expression remain unknown (Vorwerk et al. 2000; Sala-Torra et al. 2007; Boag et al. 2007). Before analysing global gene expression data and investigating clinical correlates in the PMH cohort, *CTGF* expression levels obtained by microarray were validated in cryopreserved pre-B ALL specimens by qRT-PCR. These validation experiments revealed that there was a significant correlation between microarray and qRT-PCR methods for detection of *CTGF* mRNA, confirming that the global gene expression data obtained by Boag and colleagues provided an accurate representation of *CTGF* gene expression in the PMH cohort (Boag et al. 2007).

Analysis of *CTGF* mRNA levels by qRT-PCR revealed that those primary samples that were not grouped into one of the common cytogenetic groups exhibited very low, almost undetectable *CTGF* expression. These specimens had previously been classified as *CTGF*^{low/neg} by array, based on *CTGF* probeset signals that were close to or the same as background levels. However, the extreme sensitivity of qRT-PCR confirmed that there was low level *CTGF* expression in almost all of these primary specimens. The level of *CTGF* mRNA that was detected was so low as to be unlikely to result in translation of CTGF protein, however this finding nonetheless suggests that the *CTGF* locus was in a permissive transcriptional state in these pre-B ALL specimens. These data infer that biological changes associated with *BCR-ABL1*, *TEL-AML1*, *MLL* disease and hyperdiploidy may be important in actively promoting overexpression of *CTGF*, although the underlying factors involved remain undefined. Recurrent oncogenic mutations and genomic rearrangements are common in pre-B ALL (Mullighan et al. 2007; Pui 2006), and if such genomic lesions affect the *CTGF* locus at 6q23.1 they may enhance *CTGF* gene expression. Extensive genome-wide analysis of genetic alterations in pre-B ALL have thus far failed to identify any recurrent alterations at the *CTGF* locus (Mullighan et al. 2007; Kuiper et al. 2007b), however a more focussed investigation of this region is warranted.

3.3.1 Clinical correlates of *CTGF*

The vast range of *CTGF* expression observed in the PMH cohort prompted the investigation into associations between clinical features present at the time of diagnosis and *CTGF* gene expression in patient specimens. When patients were stratified based on clinical features there was a significant association between lymphadenopathy and lower *CTGF* mRNA expression in patient specimens ($p=0.036$). There was also a similar trend with splenomegaly, however this did not reach significance. High *CTGF* expression has been positively associated with lymph node metastasis in squamous cell carcinomas of the neck (Li et al. 2007), while low *CTGF* expression is associated with lymph node metastasis in colorectal cancer (Lin et al. 2005). Thus the relationship between *CTGF* expression and lymph node metastasis is controversial and is likely to depend heavily upon tumour type and cell of origin. Validation of the association between lymphadenopathy and low *CTGF* expression in pre-B ALL should be performed in an independent patient cohort, however no such patient data was available during this study.

Details from unpublished experiments performed in the laboratory of Prof Ursula Kees imply a role for CTGF in enhancing adhesion of pre-B ALL cells to stromal elements. This finding coupled with the high incidence of bone marrow fibrosis in ALL (Kundel et al. 1964; Wallis & Reid 1989) and correlation between bone marrow fibrosis and high MRD after induction therapy (Noren-Nystrom et al. 2008), suggests that fibrotic or reactive stroma in the bone marrow microenvironment may represent a preferred sanctuary site for leukaemic blasts compared to peripheral organs. It is conceivable that aberrant CTGF secretion within the bone marrow space may result in the formation of reactive stroma, thus promoting homing of leukaemic cells to the bone marrow, rather than peripheral organs. Future functional studies focussing on the effect of CTGF-induced alterations in bone marrow stroma are expected to test such hypotheses and may confirm a role for CTGF in attracting leukaemic cells to this microenvironment.

3.3.2 *CTGF* expression is a prognostic biomarker

The prognostic importance of *CTGF* expression in pre-B ALL has been demonstrated in both adult and paediatric pre-B ALL, with *CTGF* expression correlating with low RFS (Sala-Torra et al. 2007; Kang et al. 2010). Analysis of the role of *CTGF* as a biomarker in the PMH cohort confirmed an association between *CTGF* expression and a reduced RFS. While the difference in RFS between *CTGF*^{pos} and *CTGF*^{low/neg} groups did not reach statistical significance, the present findings support those made by Kang and colleagues that *CTGF* expression can be predictive of relapse in paediatric pre-B ALL (Kang et al. 2010). The striking associations observed in the Kang and Sala-Torra studies were not observed in the PMH cohort, and this may be due to differences in treatment protocols, as well as differences in the molecular types represented in the various cohorts.

While both the Kang and PMH cohorts were comprised of paediatric pre-B ALL patients, there were significant differences in the composition of these cohorts. The Kang cohort was comprised of high-risk patients as defined by age > 10yrs, WBC > 5×10^4 / μ l, and the absence of low-risk (Hyperdiploidy, *ETV6-RUNX1*) or very high-risk (hypodiploid, *BCR-ABL1*) disease (Kang et al. 2010), whereas the PMH cohort was comprised predominantly of standard-risk patients at a ratio of 3:1 (standard:high). Furthermore, the PMH cohort is comprised predominantly of younger patients with 73% aged between 1 and 10 years, and this age group has a favourable outcome compared to those over 10 years of age (Pieters & Carroll 2010). It is unsurprising therefore that the ability of *CTGF* mRNA to independently predict relapse was more limited in the PMH cohort, as the majority of patients in this cohort were predicted to have a favourable outcome based on the assessment of risk-features alone. Analysis of clinical features associated with *CTGF* expression would be useful in the Kang cohort which consists of a highly specific subtype of ALL, however this data was not made publically available (Kang et al. 2010). Analysis of such a discreet prognostic group may infer an association between *CTGF* expression and biological features and suggest a functional role for *CTGF* in pre-B ALL.

3.3.3 Gene expression correlates of *CTGF*

In order to investigate regulation of the *CTGF* locus in pre-B ALL, interrogation of gene expression profiles correlated with that of *CTGF* was undertaken in the PMH cohort (Boag et al. 2007). Two other publically available pre-B ALL microarray cohorts were also included to ensure robust analysis (Ross et al. 2003; Kang et al. 2010). Identification of genes highly correlated with *CTGF* expression was expected to identify common transcriptional networks (Mansson et al. 2004). The five genes that demonstrated significant correlation with *CTGF* expression in all three cohorts (*SOCS2*, *MEF2C*, *ADD3*, *GSN*, and *DPYSL2*) have disparate functions, and while one of these genes, *MEF2C* encodes a transcription factor, direct involvement of MEF2C protein was ruled out due to the lack of predicted MEF2C binding sites in the 5' flanking sequence of these *CTGF*-correlated genes.

Analysis of the 5' flanking sequence of *CTGF* and the five *CTGF*-correlated genes revealed predicted binding sites for the Ikaros family of transcription factors in all 6 sequences. At least one HELIOS binding site was identified in the 5' flanking regions of all six genes, and binding motifs for IKAROS in all but the *SOCS2* flanking sequence. These predicted sites were in agreement with published IKAROS and HELIOS binding matrices (Molnar & Georgopoulos 1994). The first four members of the Ikaros family, IKAROS, HELIOS, AIOLOS and EOS, have similar DNA binding preferences through recognition of a core GGGA[A/T] pentamer motif (Molnar & Georgopoulos 1994; Perdomo et al. 2000). The fifth member; PEGASUS has very different requirements, preferring an 8pb GNNGTGNG sequence motif (Perdomo et al. 2000).

3.3.4 Ikaros family of transcriptional regulators

As discussed in Chapter 1, Ikaros proteins are sequence specific transcription factors required for normal haemopoiesis and lymphoid specification, that can both repress and potentiate gene expression through interactions with DNA and chromatin modifying complexes (Koipally et al. 1999; Kim et al. 1999; Cobb & Smale 2005). Ikaros proteins function as homodimers, but can also form heterodimers with other Ikaros family

members (Rebollo & Schmitt 2003). Mutations or in-frame deletions affecting the N-terminal zinc finger DNA-binding domain can result in the synthesis of dominant negative (DN) isoforms that bind to canonical IKAROS (or other Ikaros family proteins) and inhibit the ability of the protein dimer to bind to DNA target sequences (Sun et al. 1996). Forced expression of DN IKAROS mutants immortalises murine haemopoietic precursor cells (Ruiz et al. 2008) and loss of function mutations affecting Ikaros proteins are associated with haematological malignancies including AML, pre-B ALL and T-cell leukaemia (Yagi et al. 2002; Mullighan et al. 2007; Tabayashi et al. 2007).

Recent reports have identified deletion of IKAROS or expression of DN isoforms as a powerful indicator of relapse and a poor prognosis in high-risk pre-B ALL (Kuiper et al. 2010; Kang et al. 2010; Mullighan et al. 2008; Mullighan et al. 2009). In the present study, the relative expression levels of Ikaros family genes remained constant between $CTGF^{\text{pos}}$ and $CTGF^{\text{low/neg}}$ specimens in the PMH cohort, thus gross changes in expression of Ikaros family genes cannot account for changes in $CTGF$ expression. Analysis of the Mullighan cohort of high-risk, paediatric, pre-B ALL patients (Mullighan et al. 2009), revealed that $CTGF$ expression was significantly higher in those high-risk patients that harboured deletion or mutation of the *IKAROS* gene, compared to those with wild-type *IKAROS*. Furthermore, analysis of a sub-set of specimens harbouring genetic lesions in critical B-cell developmental genes revealed a highly significant association between *IKAROS* alterations and high $CTGF$ expression.

These data provide independent evidence to support the hypothesis that Ikaros family proteins are linked to aberrant $CTGF$ gene expression. A more detailed analysis of the role of the Ikaros family of transcriptional regulators in the context of $CTGF$ expression is warranted. Furthermore, the focus of any future study should be extended to include the first three members of the Ikaros family of proteins (*IKAROS*, *AIOLOS* and *HELIOS*), because of the similarities in their DNA binding preferences (Molnar & Georgopoulos 1994; Rebollo & Schmitt 2003), their ability to interact with and modulate the function of each other (Sun et al. 1996), as well as the presence of predicted HELIOS binding motifs in the 5' flanking sequence of all $CTGF$ -correlated

genes investigated. Such a study may uncover as-yet unidentified DN forms of other Ikaros family proteins that are of importance in modulating IKAROS function.

3.3.6 Conclusions

In summary, the findings reported in this chapter suggest that the *CTGF* locus exists in a transcriptionally permissive state in the majority of pre-B ALL specimens. A panel of five genes displayed significant correlation with *CTGF* expression in three independent patient cohorts and these five genes, as well as *CTGF* all contain predicted binding sites for the Ikaros family transcription factors IKAROS and HELIOS in their 5' flanking region. *IKAROS* mutations are significantly associated with increased *CTGF* expression in high-risk pre-B ALL, particularly in those patients with lesions affecting B-cell developmental pathways. Analysis of RFS survival in the PMH cohort demonstrated a lower RFS in patients exhibiting high *CTGF* expression, confirming findings by others in adult and paediatric ALL (Sala-Torra et al. 2007; Kang et al. 2010).

Chapter 4

Sequence and Structure of the *CTGF* Locus in Pre-B ALL

Chapter 4

Sequence and Structure of the *CTGF* Locus in Pre-B ALL

4.1 Introduction

CTGF is overexpressed in a high proportion of pre-B ALL tumours (Boag et al. 2007; Sala-Torra et al. 2007), however the mechanisms promoting its aberrant expression are unclear. The answer to this question may lie in DNA mutations or genomic rearrangements affecting the *CTGF* locus. Discrete DNA mutations can alter the ability of nuclear factors to bind to DNA and regulate gene expression. Mutation within the *CTGF* promoter or 5' flanking sequence could result in enhanced expression by disrupting sequences normally bound by repressive transcription factors such as the putative Ikaros family binding site identified in the previous chapter (Laurila & Lahdesmaki 2009; Blesa et al. 2008). Alternatively, mutation may create a new transcription factor binding site, as seen in the mutational activation of the survivin (*BIRC5*) promoter in some breast cancers (Boidot et al. 2010) or aberrant activation of superoxide dismutase 2 (*SOD2*) promoter in a broad range of carcinomas (Xu et al. 2008). As well as mutations involving the 5' flanking sequence of *CTGF*, disruption of regulatory elements in the 3' UTR may also serve to enhance *CTGF* expression by disrupting post-transcriptional, miRNA-mediated repression. Altered patterns of gene expression have been attributed to changes in miRNA levels in numerous cancers including leukaemia (Tanaka et al. 2009; Agueli et al. 2010; Volinia et al. 2010). Mutation of miRNA binding sites may result in altered *CTGF* expression or mRNA stability in pre-B ALL cells.

A feature of the genomic landscape of leukaemia is the existence of recurrent genomic rearrangements that can influence gene expression or create oncogenic fusion proteins (Pui 2006). Whole genome scans can now be performed on large patient cohorts using high resolution single nucleotide polymorphism (SNP) arrays or comparative genome hybridisation (CGH) to identify regions frequently affected by copy number alterations (CNAs) in the cancer genome. These studies have recently highlighted genetic loci that

are frequently deleted in paediatric pre-B ALL including *PAX5*, *EBF* and *IKAROS* (Mullighan et al. 2007; Harvey et al. 2010; Kuiper et al. 2007b). The long arm of chromosome 6 is frequently altered in pre-B ALL, but while the most common alterations are deletions, these events rarely affect the *CTGF* locus at 6q23.1 (Thelander et al. 2008; Mullighan et al. 2007). Nonetheless, the observed genomic instability of 6q might result in structural alterations promoting *CTGF* expression that may have been overlooked in genome wide scans. These structural alterations may exist as discreet CNAs or as complex translocations involving the *CTGF* locus. Analysis of *CTGF* expression in the PMH cohort demonstrated that *CTGF* expression was highest in leukaemic blasts containing a recurrent cytogenetic rearrangement in pre-B ALL, with the exception of those harbouring the *TCF3-PBX1* rearrangement (Boag et al. 2007). This indicates that genome instability is possibly associated with transcriptional activation of *CTGF*. Therefore the structure of the *CTGF* locus at 6q23.1 should be examined to establish whether such alterations affect this region directly.

A comprehensive assessment of the DNA sequence and gross genomic structure of the *CTGF* locus was undertaken to identify genomic lesions associated with *CTGF* overexpression. The specific aims of the experiments presented in this chapter were as follows: 1) To characterise *CTGF* expression in a panel of B-lineage ALL cell line models, 2) to examine these ALL cell lines for promoter and 3'UTR mutations that may contribute to *CTGF* overexpression, and 3) to evaluate the *CTGF* genomic locus for structural rearrangements or copy number alterations.

4.2 Results

4.2.1 Cell line models of paediatric B-lineage ALL

Cell lines represent a valuable tool for the molecular modelling of human diseases, as they provide an inexhaustible supply of proteins and nucleic acids for analysis of tumour biology, as well as facilitating functional studies *in vitro*. Numerous B-lineage ALL cell lines have been generated from primary samples in the laboratory of Prof. Kees. These

cell lines were derived by serial passage of paediatric patient specimens *in vitro* and experiments have confirmed they mirror critical features of the primary patient specimens from which they were derived (Kees et al. 2003; Beesley et al. 2006). The cell lines that were chosen as a model for B-lineage ALL in this study are shown in Table 4.1. They include three pre-B ALL, two B-ALL and two infant-ALL cell lines with B-lineage features. All of the cell lines have complex karyotypes and all harbour genetic features known to contribute to leukaemogenesis, namely; *TEL-AML1*, *TCF3-PBX1* and *MYC-IgH* translocations as well as rearrangements involving the *MLL* locus at 11q23. The *CTGF* locus is located on the long arm of chromosome 6 at cytoband 6q23.1 and there have been no alterations of this locus observed in any of these cell lines to date.

Table 4.1 Features of B-lineage ALL cell lines

Cell Line	Subtype [§]	D/R [‡]	Karyotype	Genetic Feature	Refs
PER-145	pre-B	R	45,XY, der(3)t(3;?)(q12;?), der(4)t(4;?)(p15.2;?), t(5;17)(q15;p13), der(7)t(7;8)(p13-14;q13-21), -8,?t(9;20;9)(p13;p12;q34), der(12)t(12;?)(p13;?)	<i>TEL-AML1</i>	(Kees 1987)
PER-278	pre-B	D	46,XY,der(9)t(1;9)(q23;p13),der(19)t(1;19)(q23;p13)	<i>TCF3-PBX1</i>	(Kees et al. 1990)
PER-371	pre-B	D	46,XY, der(16)t(1;16)(q2?1;p13), der(19)t(1;19)(q?13;p13) 46,X,-Y,+?der(1)t(Y;1)(q12;?q21), add(11)(q21), der(19)t(1;19)	<i>TCF3-PBX1</i>	(Kees et al. 2003)
PER-377	B-ALL	R	46,XY, t(2;13)(p12;q34), del(7)(q11q21), ?inv(14)(q11q24), der(17)t(8;17)(q11;p11)	cryptic <i>MLL-AF9</i>	(Kees et al. 1995; Whitman et al. 2001)
PER-495	B-ALL	R	46,XY,t(8;14)(q24;q32)	<i>MYC-IgH</i>	(Kees et al. 2003)
PER-485	infant-ALL	R	47,XX, der(4)t(4;11)(q21;q23) add(4)(p16),_6, del(7)(p14), add(8)(q24.3), der(9)inv(9)(p11q12) del(9)(p24), der(11)t(4;11)(q21;q23)	<i>MLL</i> rearranged	(Kees et al. 2003)
PER-490	infant-ALL	D	46,XX,t(4;11)(q21;q23) 46,XX,t(4;11),dup(1)(q12q44) 46,XX,t(4;11),der(2)t(1;2)(q12;q37)	<i>MLL</i> rearranged	(Kees et al. 2003)

§ Subtype: Classification of ALL specimens by immunophenotype.

‡ Specimens obtained at diagnosis (D) or relapse (R)

4.2.2 *CTGF* expression in B-lineage ALL cell lines

To gain an insight into the range of *CTGF* mRNA levels in these cell lines, expression was measured in seven B-lineage ALL cell lines and one commercial T-ALL cell line (JURKAT) by qRT-PCR. Those cell lines that expressed detectable levels of *CTGF* mRNA were classed as *CTGF*^{high} or *CTGF*^{low}. As shown in Figure 4.1, PER-377 expressed *CTGF* mRNA at levels approximately five and a half times that of the next highest cell line, PER-145 and was classed as *CTGF*^{high}. PER-145, PER-278 and PER-371 expressed *CTGF* at more modest levels and were classed as *CTGF*^{low}. The remaining cell lines; PER-485, PER-490 and PER-495, as well as the T-ALL cell line JURKAT, were negative for *CTGF* mRNA. All of these cell lines have been analysed previously by microarray and the *CTGF* expression data generated by qRT-PCR were in good agreement with this data.

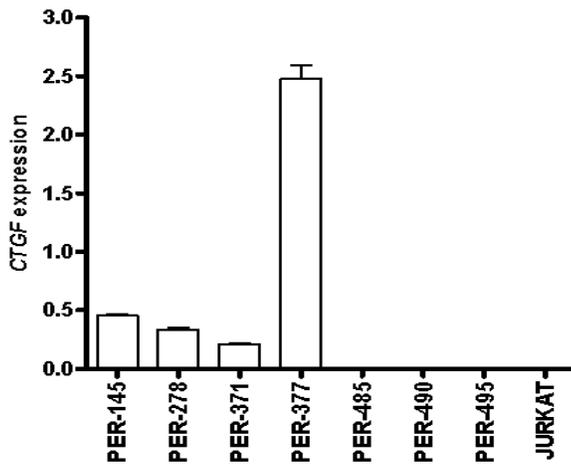


Figure 4.1 *CTGF* gene expression in ALL cell lines

CTGF expression normalized to *ACTB* was measured by qRT-PCR in seven B-lineage ALL and one T-ALL cell line (JURKAT). Error bars represent the standard error of the mean of three technical replicates.

4.2.2.1 CTGF is secreted by PER-377

CTGF is a secreted protein and may have important extracellular roles in signalling between pre-B ALL cells and the bone marrow microenvironment. B-lineage ALL cell lines were examined for CTGF protein synthesis by immunoblotting using two antibodies outlined in Figure 4.2A. These were the L20 antibody which recognised an epitope encompassing the C-terminal region of module 1 and the N-terminal region of module 2 (amino acids 150 to 200) of the CTGF protein, and the polyclonal Rabbit-A antibody which recognised the C-terminus of the CTGF protein between amino acids 247 and 260 (Brigstock et al. 1997). Whole protein extracts (75µg) from the three cell lines that exhibited the highest *CTGF* mRNA levels; PER-377, PER-145 and PER-278 were examined by immunoblotting using the L20 antibody with the T-ALL cell line JURKAT serving as a negative control. Recombinant human CTGF (rhCTGF), which was generously provided by Prof. David Brigstock, was used as a positive control.

Full-length CTGF (38kDa) was detected in extracts from PER-377 as shown in Figure 4.2B, which was classed as *CTGF*^{high} by qRT-PCR. No CTGF protein was detected in the *CTGF*^{low} cell lines PER-145 or PER-278. CTGF protein may be produced in these cells below the level of detection by immunoblotting, or alternatively the protein may be rapidly secreted making it difficult to detect in whole protein extracts. Efforts were therefore directed towards detecting secreted CTGF in cell culture supernatants. Conditioned cell culture medium (CM) from the *CTGF*^{high} cell line PER-377 was examined using the Rabbit-A antibody which was a gift from Prof. David Brigstock and has been used previously to detect CTGF proteins in biological fluids (Brigstock et al. 1997). Cells were seeded in 1ml of culture medium, containing 20% fetal calf serum (FCS) denoted as A+20, at a density of 10^6 cells/ml and grown for 24 hrs, after which time CM was collected and pooled for analysis.

Immunoblotting of 50µl of unprocessed (Raw) CM or untreated culture medium (A+20), is shown in lanes 1 and 2 of Figure 4.2C. These lanes appeared overloaded and there was excessive cross-reactivity with the Rabbit-A antibody resulting in a failure to adequately detect secreted CTGF. This cross-reactivity was attributed to the high levels

of FCS used to culture the cell lines, as FCS contains a high concentration of bovine immunoglobulins. To overcome this, a heparin-affinity purification method was developed using heparin-conjugated agarose beads. Briefly, CM harvested at 24 hrs was pooled to a volume of 5ml, then incubated with heparin-agarose beads for 24 hrs. After washing beads, the bound protein was eluted directly into 200 μ l of hot SDS-PAGE sample buffer. The CTGF protein contains a heparin binding motif in the CT domain and enriching supernatants by this method improved the detection of CTGF from conditioned medium greatly as shown in lanes 3 and 4 of Figure 4.2C.

This approach was used to screen heparin-enriched CM from eight ALL cell lines using the L20 antibody, as immunoblots performed with this antibody were of superior quality compared to those using the Rabbit-A antibody. Nevertheless, when the entire panel of cell lines was screened for secreted CTGF (shown in Figure 4.2D), only the *CTGF*^{high} cell line PER-377 secreted detectable CTGF, evident as a 36-38kDa doublet which represented the differentially glycosylated forms of the full length protein (Ball et al. 2003a). From these results it was clear that *CTGF* mRNA levels did not reflect protein levels *in vitro*. This disparity may simply be due to the high level of CTGF mRNA expression in PER-377 compared to the *CTGF*^{low} cell lines, however post-transcriptional mechanisms may also be involved in regulating CTGF protein synthesis in these cell lines.

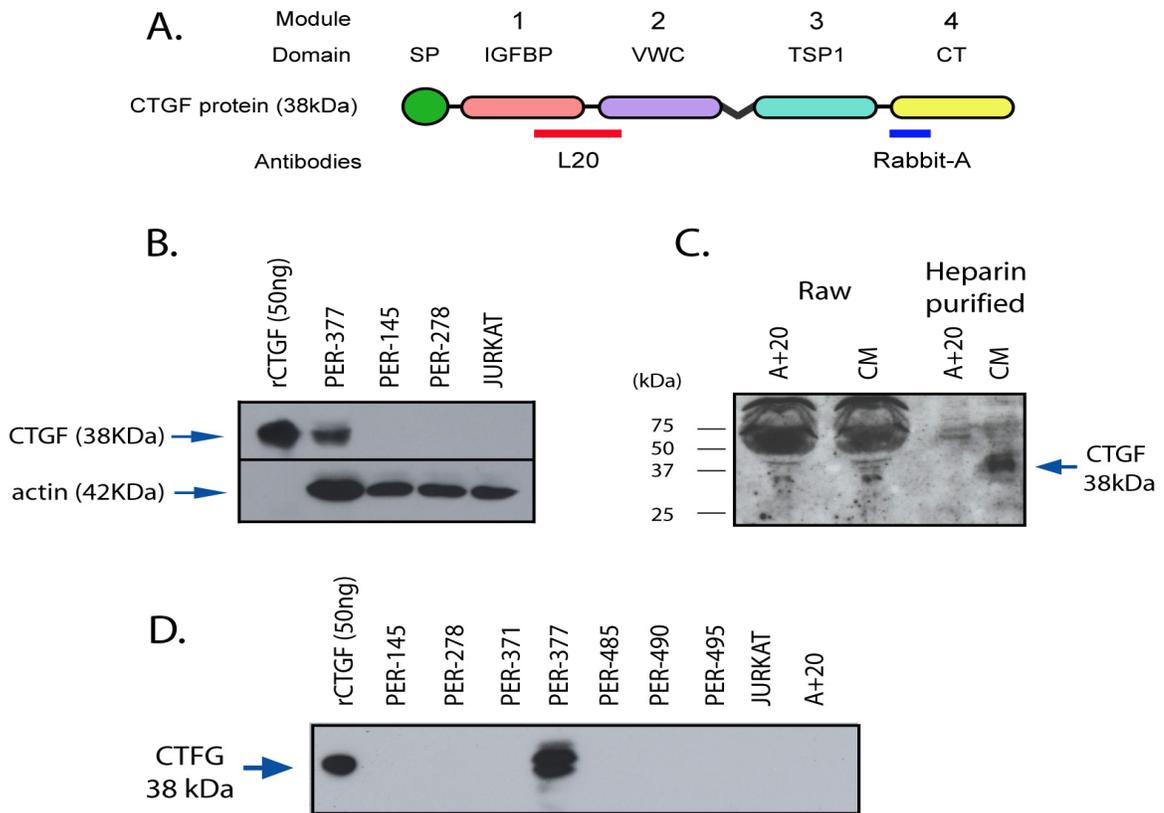


Figure 4.2 Detection of CTGF protein by immunoblotting

A. Structure of the CTGF protein and epitopes targeted by the L20 and Rabbit-A antibodies. **B.** Immunoblotting of 75µg of total protein performed with the L20 antibody. 50ng of Recombinant CTGF (rCTGF) was used as a positive control. **C.** Immunoblotting of 50µl of A+20 or CM from PER-377 without heparin purification (Raw) or after heparin purification using the Rabbit-A antibody. **D.** Detection of CTGF protein in 50 µl of heparin-purified CM from seven B-lineage ALL and one T-ALL cell line or control media (A+20) using the L20 antibody.

4.2.3 Analysis of *CTGF* 5' flanking sequence

Sequencing of the *CTGF* promoter and 5' flanking sequence was undertaken to determine if promoter mutations may contribute to dysregulation of *CTGF* expression in ALL. Three cell lines were chosen as appropriate models with which to screen the *CTGF* promoter for mutations. These were the *CTGF*^{high} cell line PER-377 and the

CTGF^{low} cell lines PER-145 and PER-278. Sequencing of the *CTGF* promoter was undertaken using three separate overlapping PCR amplicons. Figure 4.3 outlines the *CTGF* promoter and 5' flanking sequence. Validated transcription factor-binding sites are highlighted, as well as the predicted binding site for Ikaros family proteins identified in Chapter 3. Placement of primers used to amplify regions of interest and sequence PCR products are also shown. PCR products were gel extracted, purified and sequenced directly in both forward and reverse orientation and aligned to the human chromosome 6 reference sequence.

Sequencing provided complete coverage of the *CTGF* promoter in all three cell lines and high quality sequencing reads were generated in both forward and reverse orientations. The only observed deviation from the reference sequence was found in the PER-278 cell line. This cell lines was heterozygous for the rs6918698 C>G SNP at position -739 as shown in Figure 4.4A. The C>G transversion is thought to promote *CTGF* expression by disrupting repression by the SP3 nuclear factor (Fonseca et al. 2007). To determine if this polymorphism may have other functional impacts, analysis of the flanking sequence containing either the C or G allele was performed using the PROMO transcription factor binding prediction program (Farre et al. 2003; Messeguer et al. 2002) and the TFSEARCH program applied in Chapter 3. Putative transcription factor-binding sites are shown in Figure 4.4. Despite the findings by Fonseca and colleagues in relation to Sp3-mediated repression of *CTGF* expression in airway fibroblasts, neither the C nor G allele was predicted to facilitate transcription factor binding by PROMO and TRANSFAC. Nevertheless, this site has been functionally validated as a binding site for SP3 in airway fibroblasts (Fonseca et al. 2007). Further investigation into the incidence of the rs6918698 -739G genotype in *CTGF* positive specimens may be warranted.

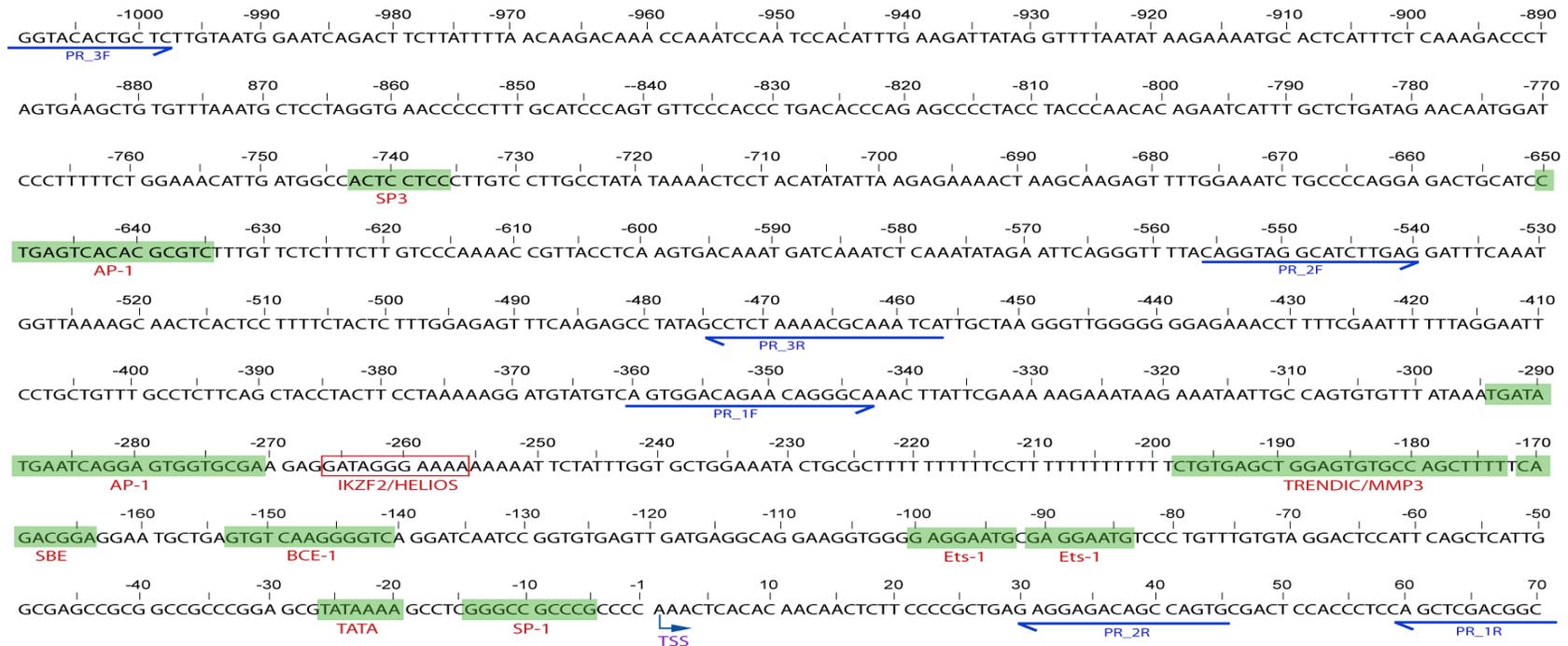


Figure 1.3 Transcription factor binding sites at the *CTGF* promoter

Primers used to generate the three sequencing amplicons (PR_1-3) from the *CTGF* promoter (-1009 to +70) are indicated by blue arrows. Validated transcription factor binding sites (green boxes) include two AP-1 sites (Yu et al. 2009), transcriptional enhancer dominant in chondrocytes (TRENDIC) which is bound by MMP3 (Eguchi et al. 2007; Eguchi et al. 2008), an SP3 binding site (Fonseca et al. 2007), a Smad binding element (SBE) (Arnott et al. 2008), basal control element (BCE-1) (Grotendorst 1997), tandem Ets-1 binding sites (Van Beek et al. 2006), a TATA box and an SP-1 binding site (Holmes et al. 2003). In addition, the location of the predicted HELIOS binding site is indicated by the red open box. TSS: transcription start site (+1).

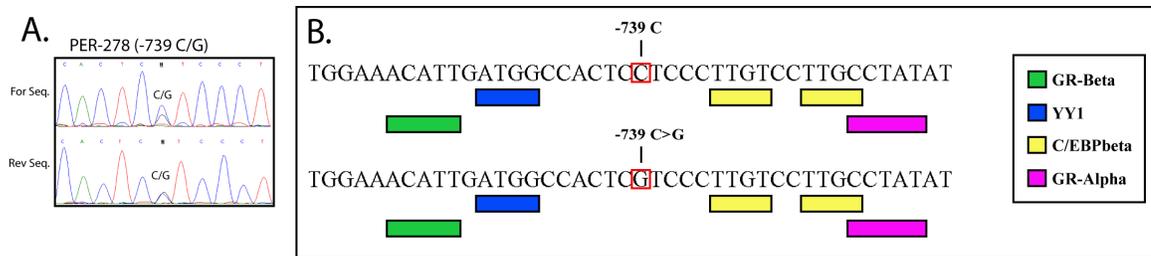


Figure 4.4 Analysis of the rs6918698 C>G SNP in PER-278

Forward and reverse sequencing of the *CTGF* promoter in PER-278 found this cell line to be heterozygous for the 6918698 C>G SNP at position -739. **B.** This SNP, highlighted by an open red box, together with flanking sequence, was examined for transcription factor binding motifs using the PROMO and TFSEARCH transcription factor binding prediction programs. Predicted TF binding motifs are indicated by coloured boxes. None of the identified motifs were associated with the rs6918698 SNP.

4.2.4 Sequencing of *CTGF* 3' UTR

The discrepancy between *CTGF* mRNA and protein levels in *CTGF*^{low} cell lines suggests some level of post-transcriptional regulation affecting the translational potential or longevity of *CTGF* mRNA. The *CTGF* 3'UTR contains several experimentally validated miRNA binding sites (Ohgawara et al. 2009; Duisters et al. 2009; Ernst et al. 2010) and mutation of miRNA binding sites may result in increased expression *via* inhibition of miRNA binding or alternatively, the introduction of a new miRNA binding site by mutagenesis may result in translational repression or increased turnover of mRNA. The miRNA target prediction software TargetScan available online at <http://www.targetscan.org/> (Friedman et al. 2009) was used to screen the *CTGF* 3'UTR for predicted miRNA binding motifs that may be of interest in addition to known binding sites. Figure 4.5 illustrates the known miRNA targets in the *CTGF* 3' UTR that have been functionally validated, including those for miR30, miR133, miR18a, as well as sites predicted by TargetScan, including binding sites for miR26a/1297, miR-132/212 and miR-19. Three PCR amplicons were generated for the 3'UTR as outlined in Figure 4.5. Primers pairs used to generate these amplicons were as follows, amplicon 1; CD_7F

& CD_7R, amplicon 2; CD_8F & CD_8R, and amplicon 3; CD_9F & CD_9R. Sequencing of these amplicons produced high quality reads in forward and reverse orientation for the PER-145, PER-278 and PER-377 cell lines. Comparing these sequences to the chromosome 6 reference sequence did not reveal any 3' UTR mutations in these cell lines. This confirmed that all known miRNA binding sites were intact in the *CTGF* 3'UTR.

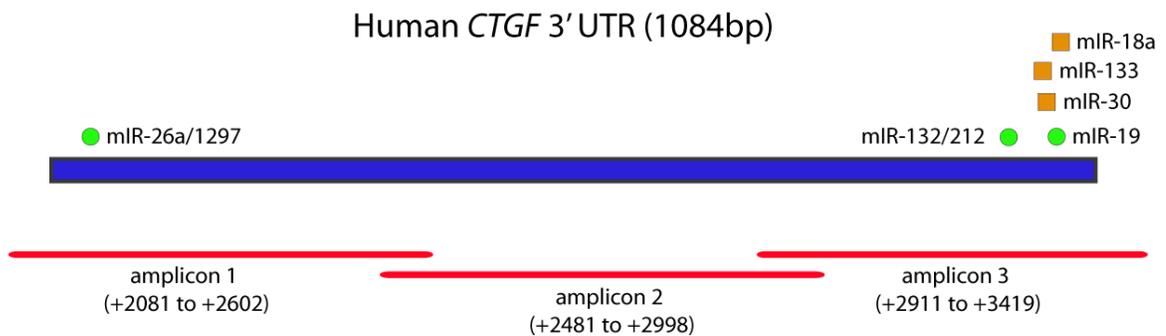


Figure 4.5 miRNA binding sites in the *CTGF* 3' UTR

The *CTGF* 3' UTR (blue) is 1084 bases in length. Predicted miRNA binding sites are indicated by green circles. Functionally validated miRNAs are shown by orange squares. Red lines (bottom) indicate overlapping PCR amplicons used to sequence the *CTGF* 3' UTR. Coordinates are relative to the *CTGF* transcription start site.

4.2.5 Structure of the *CTGF* gene locus

Besides single-base mutations, larger scale genomic rearrangements can result in activation of proto-oncogenes by juxtaposing coding regions adjacent to active promoters or similarly by moving enhancer or promoter elements to a normally silent coding region (Pui 2006). Although no 6q23 rearrangements were evident by karyotype analysis of the B-lineage ALL cell lines (shown in Table 4.1), rearrangements such as gene copy number alterations or cryptic translocations affecting the *CTGF* gene locus may influence *CTGF* gene transcription in pre-B ALL.

To examine whether the *CTGF* locus is frequently affected by genomic rearrangements in ALL, the Mitelman database of chromosomal aberrations and translocations in cancer (Mitelman et al. 2010) was screened for alterations involving 6q23.1. The results of this analysis are summarised in Table 4.2. There have not been any recorded focal amplifications of 6q23.1 in ALL, however there has been numerous cases involving deletion of regions both upstream and downstream from 6q23, indicating that this region has a relatively high frequency of rearrangement (Mitelman et al. 2010). Furthermore, data from the Mitelman database suggested that trisomy of chromosome 6 is nine-fold higher than deletion of chromosome 6 in ALL.

Table 4.2 Documented aberrations involving 6q23 in ALL

Structural alterations involving the 6q23 locus (left panel) and gross numerical alterations involving chromosome 6 (right panel) in ALL. Source: Mitelman database of chromosomal aberrations and translocations in cancer (Mitelman et al. 2010)

6q23 Alterations			Numerical Aberrations	
Band	Abnormality	Cases		
6q23	del(6)(q13q23)	18	Trisomies	
6q23	del(6)(q14q23)	4	Abnormality	Cases
6q23	del(6)(q15q23)	29	+6	892
6q23	del(6)(q16q23)	4		
6q23	del(6)(q21q23)	25	Monosomies	
6q23	del(6)(q23)	31	Abnormality	Cases
			-6	102
6q23	del(6)(q23q25)	15		
6q23	del(6)(q23q26)	2		
6q23	del(6)(q23q27)	4		

Analysis of the expression of genes proximal to the *CTGF* coding region was performed using global gene expression data from the PMH cohort described in Chapter 3, to determine if gene expression in this region was broadly deregulated or followed a similar pattern to *CTGF*. Six genes in total were analysed. These included genes on the centromeric side of the *CTGF* coding region at 6q23.1, namely A-Kinase anchor protein 7 (*AKAP7*), mediator complex subunit 23 (*MED23*) and ectonucleotide pyrophosphatase/phosphodiesterase 1 (*ENPP1*), and three genes on the telomeric side of the *CTGF* locus, namely syntaxin 7 (*STX7*), trace amine associated receptor 5 (*TAAR5*)

and *vanin3* (*VNN3*). When expression of these genes was compared between *CTGF* positive and low/negative samples, there was no significant difference in expression between the two groups, as shown in Figure 4.6. Based on these data it is unlikely that deregulation of the 6q23 region as a whole contributes to *CTGF* overexpression in pre-B ALL.

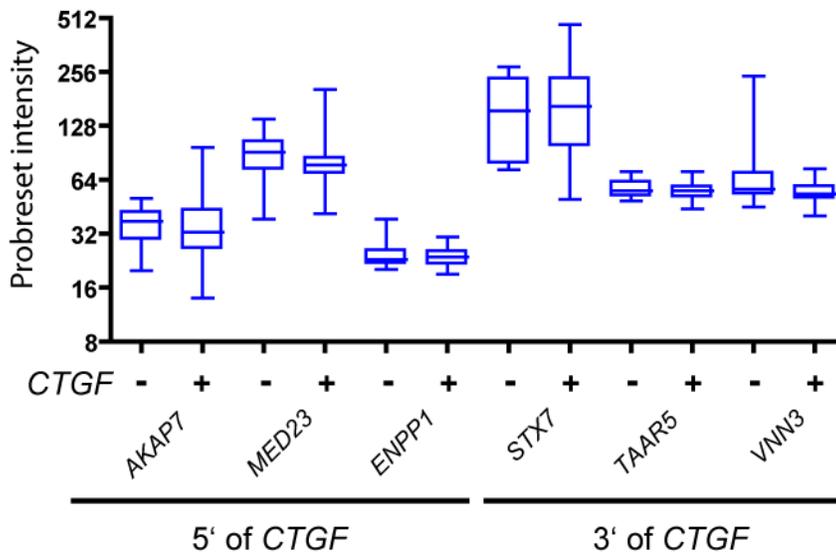


Figure 4.6 Expression of genes neighbouring the *CTGF* coding region

Expression of six genes in close proximity to the *CTGF* coding region at 6q23.1 measured in the PMH cohort by microarray. Genes were selected based on their position relative to the *CTGF* coding region. *AKAP7*; 205771_s_at, *MED23*; 218846_at and *ENPP1*; 205066_s_at are located on the centromeric side of the *CTGF* coding region. while *STX7*; 212631_at, *TAAR5*; 221459_at and *VNN3*; 220528_at are located on the telomeric side of the *CTGF* gene. For each gene, expression levels are shown in *CTGF*^{low/neg} (minus symbol, n=13) or *CTGF*^{med/high} (plus symbol, n=60) samples.

4.2.5.1 Southern blotting of the *CTGF* locus

There is evidence that internal deletions affecting the long arm of chromosome 6 occur at a relatively high frequency in pre-B ALL (Thelander et al. 2008). Genome instability may result in translocations that have profound effects on gene expression. To

investigate whether the *CTGF* locus is affected by genomic rearrangements, Southern blotting was performed on seven B-lineage ALL cell lines. The hybridisation probe and restriction sites used are outlined in Figure 4.7. Genomic DNA was digested with either Hind III or Psi I, gel fractionated and hybridised with a 4.4kb probe spanning the *CTGF* locus from -1024 to +3418 relative to the transcription start site (TSS). Digestion with Hind III was predicted to yield two fragments; 12.5kb and 2kb, capable of being hybridised by the *CTGF* probe, while digestion with Psi I produced two fragments; 2.6kb and 6.5kb. These two digests provided coverage of the *CTGF* locus from nucleotide position -9660 to +6224. Figure 4.8 shows the result of Southern blotting performed on digested genomic DNA from seven B-lineage ALL cell lines and one PBMC control sample from a healthy donor (C5). Fragments of the expected size were observed in all cases with no extraneous bands identified. Thus, there were considered to be no detectable rearrangements of the *CTGF* locus present in these samples.

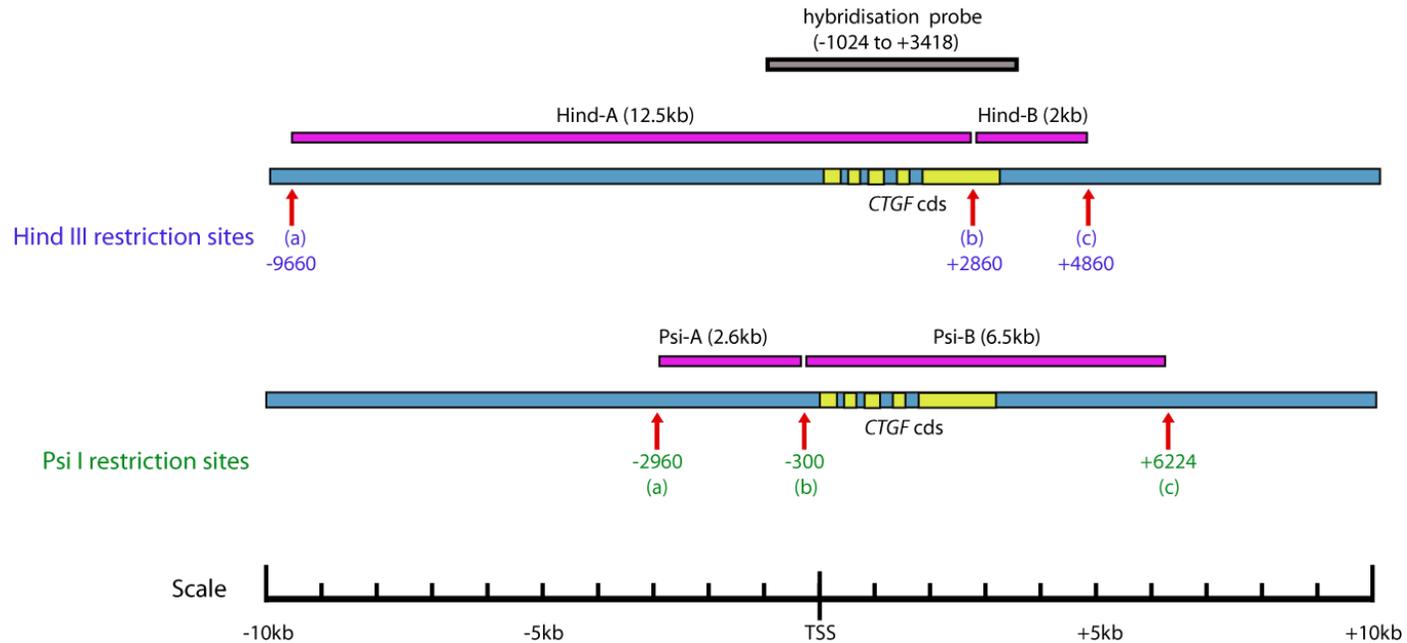


Figure 4.7 Localisation of hybridisation probe and restriction sites used for Southern blotting analysis of the *CTGF* locus
 10kb either side of the *CTGF* transcription start site (TSS) is shown above. Sequence flanking the *CTGF* coding sequence (cds) is shown in blue, while exons are indicated by yellow boxes. Scale bar at the bottom indicates distance in kilobases. Restriction sites for Hind III (top) are denoted by red arrows at -9660 (a), +2860 (b) and +4860 (c), while restriction sites for Psi I (bottom) occur at -290 (a), -300 (b) and +6224 (c). A 4.4kb *CTGF* probe (grey bar at top) spanning from -1024 to +3418 hybridises digest fragments shown as purple bars; Hind-A (12.5kb) and Hind-B (2kb) as well as two Psi I fragments; Psi-A (2.6kb) and Psi-B (6.5kb).

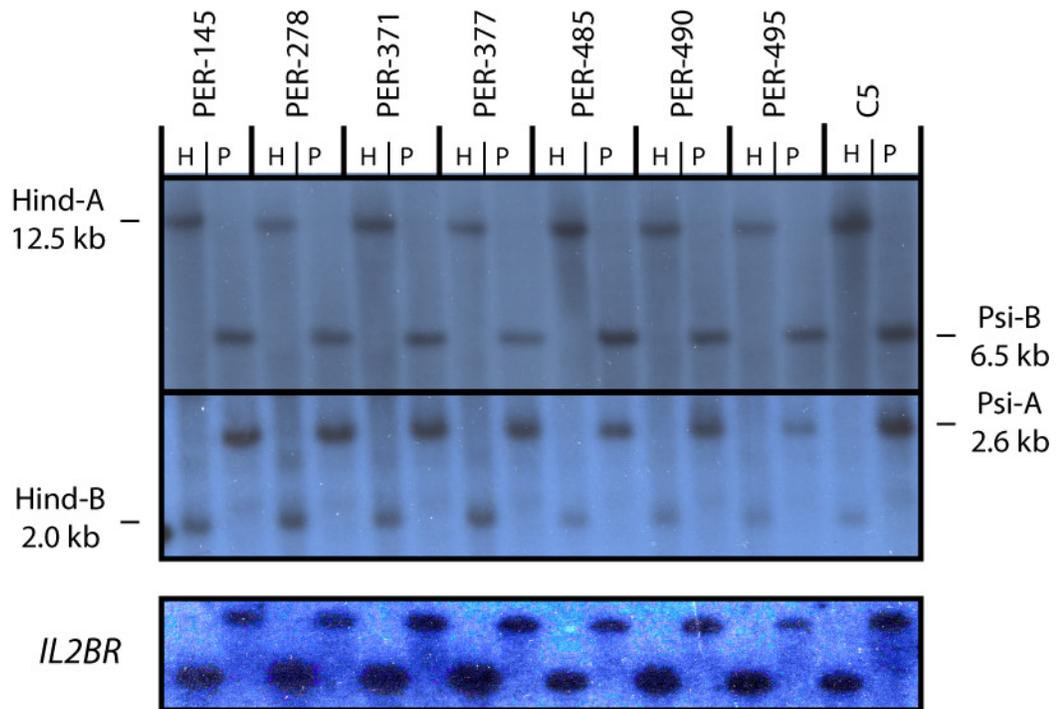


Figure 4.8 Southern blot analysis of the *CTGF* locus

Seven B-lineage ALL cell lines and PBMC gDNA from a healthy donor (C5) analysed by Southern blotting. Samples were digested with either Hind III (H) or Psi I (P) and hybridised to a cDNA probe specific for the *CTGF* locus (see Figure 4.7). Different exposures are shown for optimal resolution of large (top panel) and small fragments (middle panel). The top panel contains large fragments (Hind-A and Psi-B), while middle panel contains smaller MW fragments (Hind-B and Psi-A). Bottom panel contains samples re-probed for the *IL2BR* gene as a loading control.

4.2.5.2 Gene copy number analysis

Southern blotting did not indicate any overt rearrangements of the *CTGF* locus, however there may be CNAs outside of the region detected by Southern blotting. To further examine CNAs, a commercial qPCR gene copy number assay (Applied Biosystems) (*CTGF*:hs00517234_cn) was used to examine 7 B-lineage ALL cell lines and 17 primary pre-B ALL specimens (outlined in Chapter 3). The *CTGF* copy number assay was performed in a duplex reaction with the reference copy number assay targeting the RNase P (*RPPHI*) coding region at 14q11. Assays were run in duplex reactions in two independent experiments and data was imported into the CopyCaller software package (Applied Biosystems) for analysis. Samples were considered to contain a copy number alteration if they crossed a threshold of 3.0 indicated on the Y axis in Figure 4.9. The reference PBMC DNA samples A4 and A5 were isolated from healthy donors and used as diploid calibrator samples for these experiments. Analysis of the data using the CopyCaller software predicted alterations in *CTGF* copy number in two cell lines; PER-377 (amplification) and PER-495 (deletion).

Upon manual inspection of the raw copy number assay data, it was noted that there was significant variation in the cycle threshold (C_t) values obtained for the RNaseP control assay in cell lines, which may confound interpretation of *CTGF* copy number. Figure 4.10A shows the raw C_t values generated by both assays from the B-lineage ALL cell lines. The *CTGF* assay C_t values suggested that only PER-495 contained a genuine alteration in *CTGF* copy number, namely a deletion of one *CTGF* allele. Data obtained from the RNaseP assay suggested that PER-377 and potentially PER-490 may contain less than two copies of the RNaseP locus. Furthermore, RNaseP is located at 14q11.2 and the cytogenetics of PER-377 indicates that this cell line harbours a cryptic rearrangement on chromosome 14 between q11 and q24.

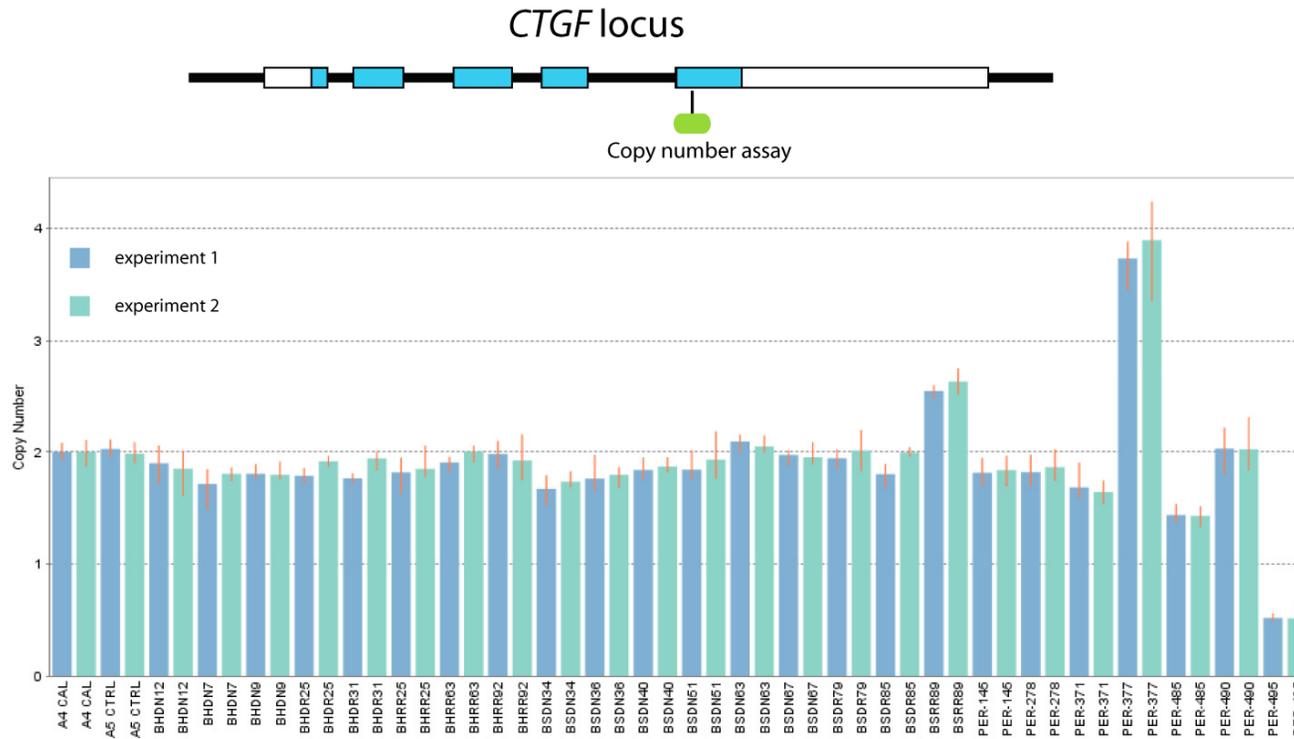


Figure 4.9 *CTGF* copy number analysis

Top; the *CTGF* copy number assay (hs00517234_cn) targets exon 5 within the *CTGF* coding region. Bottom; results of the copy number assay from 2 independent experiments normalised against the reference assay for ribonuclease P RNA component H1 (*RPPH1*). For each experiment, error bars represent the standard error of the mean of technical replicates (n=4). A4 and A5 gDNA was obtained from PBMC isolated from healthy individuals.

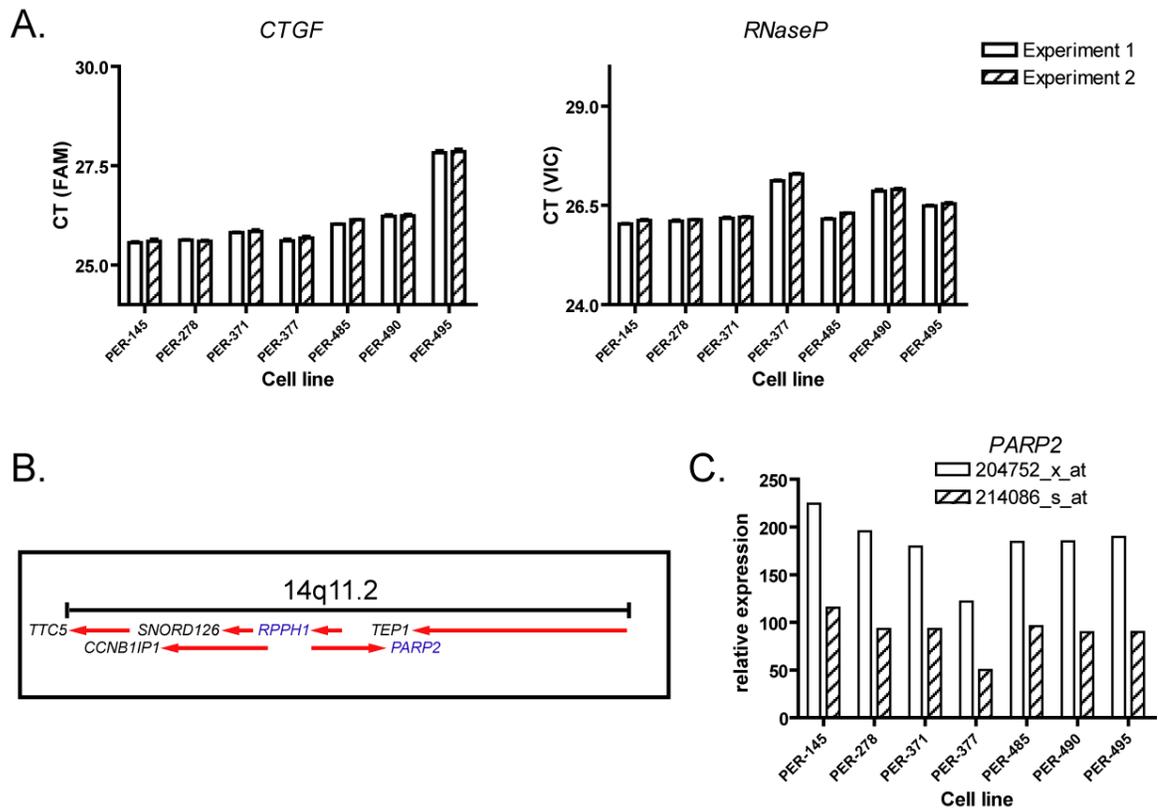


Figure 4.10 Raw gene copy number data obtained from ALL cell lines

A. Taqman copy number assay data generated from the *CTGF* assay (left) and the reference *RNaseP* assay (right). Raw (non normalized) C_t values are plotted for each of the seven B-lineage ALL cell lines. Two biological replicates (experiments 1 and 2) were conducted. Error bars represent the standard error of the mean of technical replicates ($n=4$). **B.** The *RNaseP* copy number reference assay targets the coding region of *RNaseP* (*RPPHI*) at 14q11.2. The *PARP2* gene is located on the opposite strand at this locus. **C.** *PARP2* expression measured by microarray (204742_x_at, 214086_s_at) in seven B-lineage ALL cell lines.

Gene expression at the 14q11.2 locus was examined, as loss of one allele should be reflected in changes in expression of affected genes. There was no Affymetrix probeset targeting *RNaseP* (*RPPHI*) on the U133A platform, however there is a probeset for *PARP2*, which is located in close proximity to *RPPHI* on the opposite genomic strand as shown in Figure 4.10B. The expression level of *PARP2* was determined using microarray profiles from our B-lineage ALL cell lines and it was found that there was

indeed a reduced expression of the two *PARP2* probesets (204752_x_at, 214086_s_at) in PER-377 (Figure 4.10-C), supporting the hypothesis that PER-377 harbours a deletion at the *RPPH1/PARP2* locus. Based on this finding as well as the observed differences in C_t values for the *RNaseP/RPPH1* copy number assay, the most likely explanation is that PER-377 contains two copies of *CTGF* and harbours a deletion of *RPPH1*, while PER-495 contains a focal deletion of one copy of *CTGF*. All primary specimens tested were diploid for *CTGF*.

4.3 Discussion

Alterations in DNA sequences are considered to be the earliest events underpinning neoplastic transformation of pre-cancerous cells (Greaves 2009). Recurrent chromosomal translocations and genome instability are a feature of leukaemia cells, thus determining if overexpression of *CTGF* mRNA occurs in the absence of DNA rearrangements or mutations was considered an important step in elucidating the molecular mechanisms contributing to *CTGF* deregulation. Analysis of the *CTGF* locus was carried out in a panel of seven B-lineage ALL cell lines, which had been previously evaluated for *CTGF* expression (Kees et al. 2003; Beesley et al. 2006). Further analysis of *CTGF* expression levels in these cell lines revealed that only the *CTGF*^{high} cell line PER-377 synthesised CTGF protein. This suggested that post-transcriptional regulation may repress CTGF protein synthesis in the *CTGF*^{low} cell lines. A study into the prognostic significance of *CTGF* mRNA expression in adult ALL observed a poorer outcome in patients that expressed *CTGF* mRNA at high levels in their leukaemic blasts compared to those with low or intermediate level expression (Sala-Torra et al. 2007), thus there may be a threshold below which CTGF protein synthesis is repressed post-transcriptionally. Stratifying patients in the PMH cohort by *CTGF* mRNA expression (high, medium or low) was not predictive of RFS or outcome (data not shown), thus careful immunological examination of CTGF protein expression in trephine bone marrow biopsies should be performed and coupled with qPCR analysis from matched leukaemic blasts, to determine if such a threshold exists as this may have significant prognostic relevance for pre-B ALL patients.

The 3' UTR in exon 5 of *CTGF* contains several miRNA binding sites that have been experimentally validated, as well as a number of predicted binding sites identified in the present study (Duisters et al. 2009; Ohgawara et al. 2009). Sequencing this region in three *CTGF* positive cell lines confirmed that no mutations affected the 3' UTR, thus full-length, canonical *CTGF* transcripts should be amenable to post-transcriptional regulation by miRNAs. Profiling of miRNA expression levels in pre-B ALL was considered beyond the scope of the present study, however post-transcriptional regulation by miRNAs may be responsible for the apparent lack of CTGF protein in *CTGF*^{low} cell lines. Global miRNA profiling integrated with gene expression data would provide some clues as to the role of miRNAs in regulating *CTGF* and other genes in pre-B ALL (Lionetti et al. 2009; Volinia et al. 2010). Furthermore, inhibiting specific miRNAs such as those already validated as *CTGF* regulators, in *CTGF*^{low} cell lines may confirm a role for these miRNAs in *CTGF* expression in pre-B ALL.

4.3.1 *CTGF* promoter mutation was not detected in pre-B ALL

Mutation of the *CTGF* promoter or 5' flanking sequence was considered a possible mechanism by which *CTGF* overexpression may be potentiated in pre-B ALL. Analysis of the *CTGF* locus confirmed that the promoter and 5' flanking region from pos -1009 to +70 did not contain any DNA sequence mutations. However, the PER-278 cell line harboured the rs6918698 C>G SNP at -739. This polymorphism is contained within a binding site for the ubiquitously expressed nuclear factor SP3, and transversion from C to G can disrupt SP3-mediated repression of the *CTGF* locus (Fonseca et al. 2007). Functional studies conducted by Fonseca and colleagues in primary airway fibroblasts confirmed that SP3 can bind to this site when the -739 C nucleotide is present and repress basal and thrombin-induced *CTGF* promoter activation (Fonseca et al. 2007). SP3 is a dual function transcription factor that can induce or repress gene expression through interaction with different co-factors. SP3 can undergo post-translational modification by small ubiquitin-like modifier (SUMO) proteins, and these modifications promote repression of SP3-bound loci (Stielow et al. 2008). Furthermore, SUMO proteins have been shown to recruit chromatin-modifying enzymes and DNA

methyltransferases, resulting in compact, transcriptionally silent heterochromatin and increased site-specific DNA methylation, thus SP3 may act to target or maintain epigenetic silencing at the *CTGF* locus (Stielow et al. 2010).

Genotyping studies have investigated the association of this SNP with *CTGF* overexpression in systemic sclerosis (SSc), as serum and dermal *CTGF* levels are elevated in patients with this disease (Sato et al. 2000; Igarashi et al. 1995). Some studies have identified significant associations between the rs6918698 G allele and SSc (Fonseca et al. 2007; Kawaguchi et al. 2009), however, other studies have not been able to replicate these associations in independent cohorts (Rueda et al. 2009; Gourh et al. 2008; Granel et al. 2010). While this SNP was only identified in one of the three cell lines that were sequenced, it may have a higher penetrance in primary pre-B ALL specimens. Based on the findings by Stielow and colleagues, germline transversion of this SNP to the G allele may result in a failure of SP3 to target epigenetic repression of the *CTGF* locus and subsequently the *CTGF* locus would be permissive to improper transcriptional activation (Stielow et al. 2010). Previous studies investigating frequent genetic alterations in pre-B ALL have employed genome-wide SNP-array technology (Kuiper et al. 2007b; Mullighan et al. 2007), however a probe for the the rs6918698 SNP is not contained on Affymetrix array platforms and thus was not interrogated in these studies. Future experiments are justified to undertake genotyping of this SNP in a cohort of primary pre-B ALL specimens, and to establish if there is an association between the rs6918698 SNP and *CTGF* gene expression more broadly in ALL.

4.3.2 The *CTGF* locus in not structurally altered in pre-B ALL

In addition to single base DNA mutations, genomic translocation or gene amplification may be associated with *CTGF* overexpression. Genome wide scans have identified frequent internal deletions in the long arm of chromosome 6, however the *CTGF* locus remains largely unaffected by such alterations (Thelander et al. 2008; Mullighan et al. 2007). This finding was confirmed by analysis of the Mitelman database of Chromosome Aberrations and Gene Fusions in Cancer which contains cytogenetic

information from 58,819 primary cancers (Mitelman et al. 2010). Analysis of 6q alterations in ALL revealed that areas adjacent to *CTGF* at 6q23.1 are frequently affected by deletions. This genetic instability may result in translocation or rearrangement of the *CTGF* locus, or the juxtaposition of enhancer elements from distant loci close to the *CTGF* promoter. Analysis of gene expression in the PMH cohort revealed that expression of genes adjacent to *CTGF* at 6q23 was independent of *CTGF* expression, ruling out amplification or any kind of positional effect upon 6q23 as a whole. Southern blotting confirmed that the *CTGF* locus was not rearranged in the ALL cell lines studied, and direct copy number analysis of the *CTGF* coding region by qPCR ruled out amplification of the *CTGF* coding region in seven ALL cell lines and 17 primary pre-B ALL specimens. These findings confirmed that structural alterations of the *CTGF* locus do not contribute to *CTGF* overexpression in pre-B ALL. Gene copy number analysis indicated that the *CTGF*^{high} cell line PER-377 was diploid for *CTGF*.

4.3.3 Conclusions

In summary, the experiments presented in this chapter identified a disparity between *CTGF* mRNA and protein levels suggesting that post-transcriptional regulation of *CTGF* mRNA may occur in *CTGF*^{low} cell lines. This finding highlights the need to evaluate leukaemic blasts for *CTGF* protein using immunological techniques, as qRT-PCR alone may not be sufficient to gauge the *CTGF* protein status of leukaemic cells. Analysis of the *CTGF* locus in B-lineage ALL cell lines and primary patient specimens ruled out structural alterations in the *CTGF* locus as contributing to *CTGF* expression in pre-B ALL. Furthermore, the 5' flanking sequence and 3' UTR were wild-type, confirming that promoter and 3' UTR mutation is not a mechanism by which *CTGF* expression is promoted in pre-B ALL. Finally, this study highlights the potential importance of the rs6918698 C>G SNP in the repression of the *CTGF* locus and suggests that further analysis of this SNP by genotyping in patient cohorts is warranted.

Chapter 5

Characterisation of *CTGF*

Transcription in B-lineage ALL

Chapter 5

Characterisation of *CTGF* Transcription in B-lineage ALL

5.1 Introduction

Aberrant *CTGF* gene expression in pre-B cell ALL has been identified in a number of studies as previously discussed. These findings have all been made based upon the level of *CTGF* mRNA detected in leukaemic blasts by either qRT-PCR or array based methods (Vorwerk et al. 2000; Sala-Torra et al. 2007; Boag et al. 2007). It is unclear whether *CTGF* mRNA transcribed in leukaemic cells is only represented by the canonical 2.4 kb transcript or whether there may be other *CTGF* transcripts synthesised in these cells that exhibit altered coding potential. This possibility was highlighted by findings presented in Chapter 4 showing that CTGF protein was detectable in only one of four cell lines that were positive for *CTGF* mRNA. While there were large quantitative differences in *CTGF* mRNA levels between the *CTGF*^{high} cell line PER-377 and the three *CTGF*^{low} cell lines PER-145, PER-278 and PER-371, the disparity in CTGF protein expression raises the possibility that there may be non-canonical *CTGF* transcripts synthesised in these cell lines that are either non-coding or encode CTGF protein isoforms that were not able to be detected by the antibodies used in this study.

Alternative splicing of pre-mRNA to generate transcript variants can increase protein coding diversity without increasing genome size (Hallegger et al. 2010). On the other hand, alternative splicing can also result in the synthesis of truncated mRNA transcripts that may be destined for degradation by nonsense-mediated decay (Peltz et al. 1994), or transcripts missing important regulatory elements such as the 3' UTR, thus modulating protein expression. Since original observations of alternative splicing almost 35 years ago, a great number of mRNA transcript variants have been recorded and it is now estimated that around 90% of human multi-exon genes are alternatively spliced (Mortazavi et al. 2008; Hallegger et al. 2010). Alternative splicing of *CCN* mRNAs including *CCN1*, *CCN3* and *CCN4* has been reported by several groups in the last decade (Perbal 2009). Furthermore *WNT1* (*CCN4*) transcript variants have been

observed in numerous tumour types, including gastric carcinoma, chondrosarcoma and hepatoma (Tanaka et al. 2001; Yanagita et al. 2007; Cervello et al. 2004), while alternative *CYR61* (*CCNI*) transcripts have been identified in breast cancer cell lines (Hirschfeld et al. 2009).

Despite evidence supporting alternative splicing in other CCN family members, alternative *CTGF* transcripts documented by sequencing has not been reported. However, there is some evidence for their existence in the form of truncated cDNAs in expressed sequence tag (EST) databases and a recent study examining the exome of thymic tumours (Soreq et al. 2008). In this study, the authors used Affymetrix exon arrays to identify overexpression of *CTGF* in thymoma specimens. Analysis of the signals generated by probes targeting individual *CTGF* exons indicated a proportional increase in exon 5 and a concurrent decrease in exon 2 compared to other *CTGF* exons (Soreq et al. 2008). This is compelling evidence for the existence of alternative *CTGF* transcripts in these thymic tumours. However, this study was conducted using exon arrays, and no further characterisation of *CTGF* transcripts by sequencing was obtained.

Previous characterisation of *CTGF* transcription was performed using either the Affymetrix microarray platform or qRT-PCR (Vorwerk et al. 2000; Sala-Torra et al. 2007; Boag et al. 2007; Kang et al. 2010). These technologies target discreet regions within *CTGF* mRNA and cannot indicate the size or structure of *CTGF* mRNA transcribed in cells under test. The experiments detailed in this chapter sought to examine *CTGF* mRNA transcripts expressed in B-lineage ALL cell lines, and to characterise these by sequencing any novel *CTGF* transcripts that were identified. The specific aims of the experiments outlined in this chapter were as follows: 1) to evaluate the size and range of *CTGF* mRNA transcripts expressed in ALL cell lines, 2) to characterise any suspected *CTGF* mRNA variants by sequencing, 3) to characterise the 5' and 3' ends of *CTGF* mRNA transcripts.

5.2 Results

Analysis of *CTGF* expression in a panel of B-lineage ALL cell lines in Chapter 4 (summarised in Table 5.1 below) demonstrated that the PER-377 cell line displayed the highest level of *CTGF* mRNA expression. This cell line was therefore chosen to examine *CTGF* mRNA transcripts by northern blotting. Three timepoints were examined to determine if *CTGF* expression was growth dependant. The PER-377 cell line was grown to confluency (2×10^6 cells/ml) and then split 1:1 with fresh culture medium. RNA was extracted at 24, 48 and 72 hrs and two independent northern blots were performed using two different probes designed to target distinct regions of the *CTGF* coding region (Figure 5.1A). The central probe hybridised to the 3' half of the coding region, corresponding to exons 3, 4 and 5, while the 3' probe hybridised to the exon 5 encoded *CTGF* 3' UTR in the same region as the *CTGF* probeset sequence (209101_at) on the Affymetrix human genome array U133A.

Both probes hybridised to the canonical 2.4 kb *CTGF* transcript, as well as to three novel shorter transcripts of approximately 1.8, 1.6 and 1.3 kb in size (Figure 5.1B). While these bands were of a lower intensity, they provided evidence that there were minor *CTGF* transcripts produced in PER-377 and these warranted further investigation. A densitometric analysis (Figure 5.1C) indicated that expression of canonical *CTGF* was highest at 24hrs compared to the later timepoints when normalised to the *ATP5G3* control. Consequently, this timepoint was chosen to analyse the remaining pre-B ALL cell lines by northern blotting in order to determine whether these non-canonical transcripts were also produced in other ALL cell lines.

Table 5.1 *CTGF* mRNA expression in B-lineage ALL cell lines

Expression levels are based upon qRT-PCR values reported in Chapter 4.

B-lineage ALL cell line	<i>CTGF</i> mRNA expression
PER-377	<i>CTGF</i> ^{high}
PER-145 PER-278 PER-371	<i>CTGF</i> ^{low}
PER-485 PER-490 PER-495	<i>CTGF</i> ^{neg}

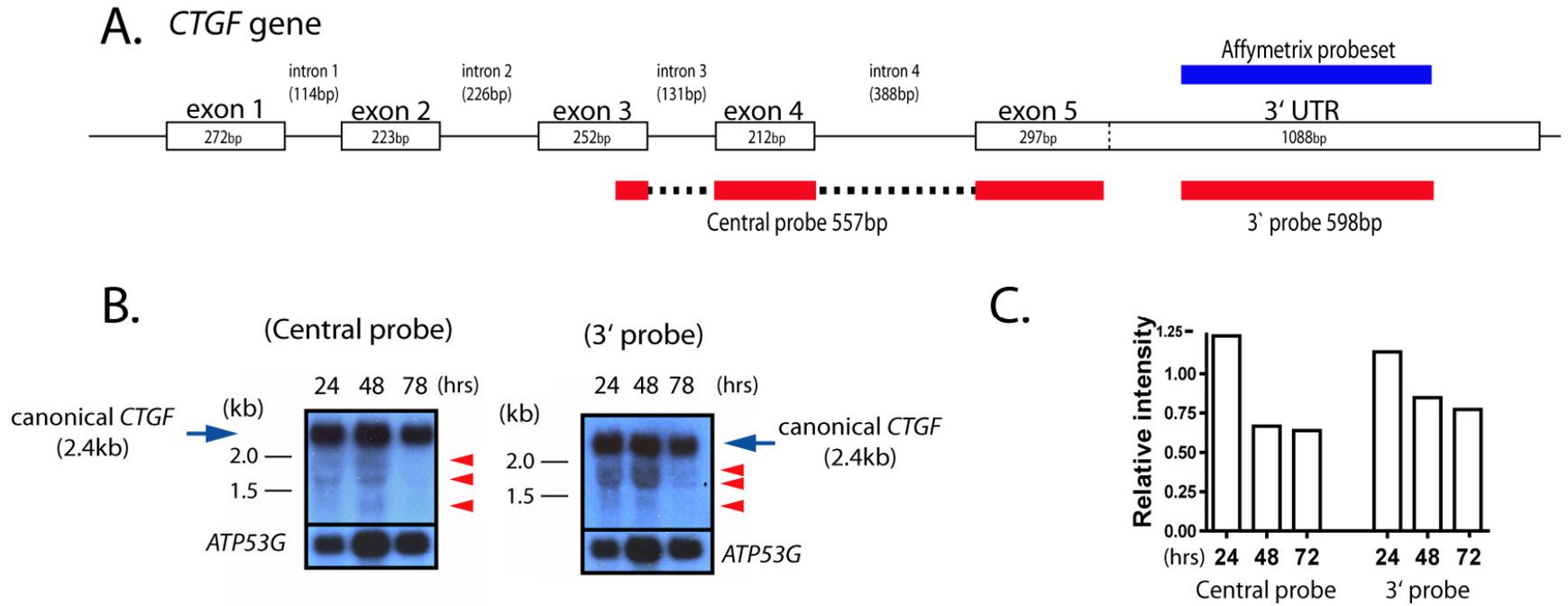


Figure 5.1 Expression of *CTGF* mRNA transcripts in the PER-377 cell line

A. Schematic representation of *CTGF* gene structure and probes used for northern blot analysis. Central and 3' probes (red bars) hybridise exons 3 to 5 or the 3' UTR. The 3' probe matches the Affymetrix *CTGF* probeset 209101_at target sequence (blue bar). **B.** Northern blots showing PER-377 RNA extracted at timepoints 24, 48 and 72 hrs. Blots were hybridised with either the central (left) or 3' probe (right). Canonical *CTGF* mRNA (2.4 kb) is indicated by the blue arrow. Red arrowheads indicate novel transcripts that were hybridised by both probes. A probe specific for *ATP53G* was used as a loading control. **C.** Relative abundance of canonical *CTGF* transcript at each timepoint calculated by densitometry using the signal for *ATP53G*.

5.2.1 Northern analysis of *CTGF* mRNA in ALL cell lines

While *CTGF* expression was much higher in PER-377 compared to the remaining cell lines utilised in this study, all cell lines were examined by northern blot to determine if the minor *CTGF* transcripts were expressed more generally in B-lineage ALL. Characterisation of *CTGF* mRNA was therefore performed on the B-lineage ALL cell lines described in Chapter 4, as well as the T-ALL cell line JURKAT which was employed as a negative control. Two independent northern blots utilising the two different *CTGF* probes were performed, as shown in Figure 5.2. The *CTGF*^{neg} PER-485 cell line was not available for analysis due to persistent RNA degradation occurring in three independent extractions. In agreement with the qRT-PCR data presented in Chapter 4, PER-377 displayed exceptionally high levels of *CTGF* mRNA compared to the other cell lines. For this reason, PER-377 is not shown in Figure 5.2A, due to high signal intensity with the central probe, but is shown hybridised by the 3' probe in Figure 5.2B.

Full-length *CTGF* was detectable in B-lineage ALL cell lines PER-145, PER-278 and PER-371 but not in PER-490, PER-495 or the T-ALL cell line JURKAT. These findings were in good agreement with the previous qRT-PCR experiments. Again, the central probe hybridised to 2.4 kb *CTGF* mRNA, as well as a shorter transcript of approximately 1.6 kb in PER-145, PER-278 and PER-371. The minor 1.3 kb transcript was also present in these cells lines, but at very low levels. When samples were examined with the 3' probe, the shorter transcripts observed in Figure 5.1, were evident only in the PER-377 cell line. This may be for technical reasons as the signal intensity of the 3' probe hybridised to the canonical transcript was less intense compared to the central probe in two independent experiments. Nonetheless, this experiment confirmed that the remaining *CTGF*^{pos} cell lines produce alternative *CTGF* transcripts and the features of these transcripts remained to be elucidated.

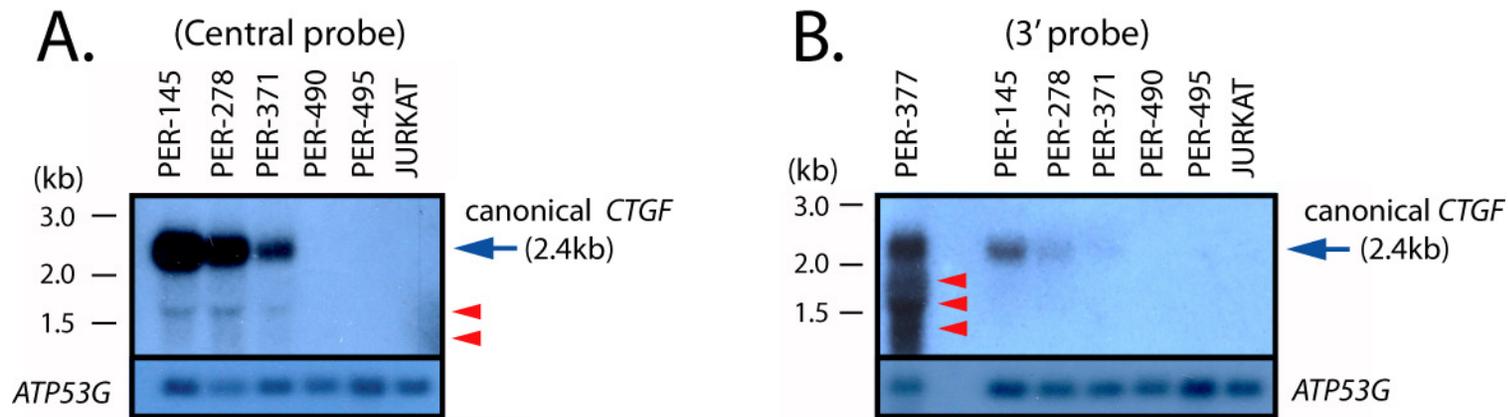


Figure 5.2 Characterisation of *CTGF* mRNA transcripts in ALL cell lines

A. Northern blot of ALL cell lines using the central probe. PER-377 was excluded from this figure due to high signal intensity. **B.** Northern blot of ALL cell lines using the 3' probe. For both blots, the canonical 2.4kb transcript is indicated by the blue arrow. Novel transcripts are indicated by red arrowheads. The T-ALL cell line JURKAT was included as a negative control. A probe designed to hybridise *ATP53G* mRNA was used as a loading control.

5.2.2 Amplification and sequencing of novel *CTGF* transcripts

Novel *CTGF* mRNA transcripts were identified by northern blotting in cell lines that were previously shown to be positive for *CTGF* expression by qRT-PCR. To examine whether these *CTGF* variants were produced by alternative splicing, a nested reverse-transcriptase PCR (RT-PCR) strategy was employed to enable their amplification and sequencing. OligodT primed cDNA was generated from *DNaseI*-treated, total cellular RNA extracts isolated from PER-377 cells in log phase growth, as they displayed the highest level of *CTGF* expression and were shown to contain these novel *CTGF* transcripts. *CTGF* cDNA was amplified from exon 1 to exon 5 and these PCR products were gel extracted, TA cloned and sequenced. Primer placement relative to the *CTGF* gene is shown in Figure 5.3A.

Somewhat surprisingly, amplification of a product corresponding to the canonical mRNA with these primers (1023 bp) was not observed. Instead, two shorter products of approximately 330 bp and 250 bp were amplified (Figure 5.3B). These PCR products were cloned and sequenced, revealing that only the larger PCR product was amplified from *CTGF* cDNA. This 331 bp product was amplified from a novel *CTGF* transcript generated by alternative splicing. Sequencing confirmed that this transcript was missing exons 2 and 3, and contained truncated forms of exons 1 and 4. Splicing of truncated exon 1 to exon 4 occurred 73 bases before the start of the *CTGF* coding region at position +207. Splicing from exon 4 to 5 adhered to canonical exon junctions. This novel cDNA was designated as Variant 1.

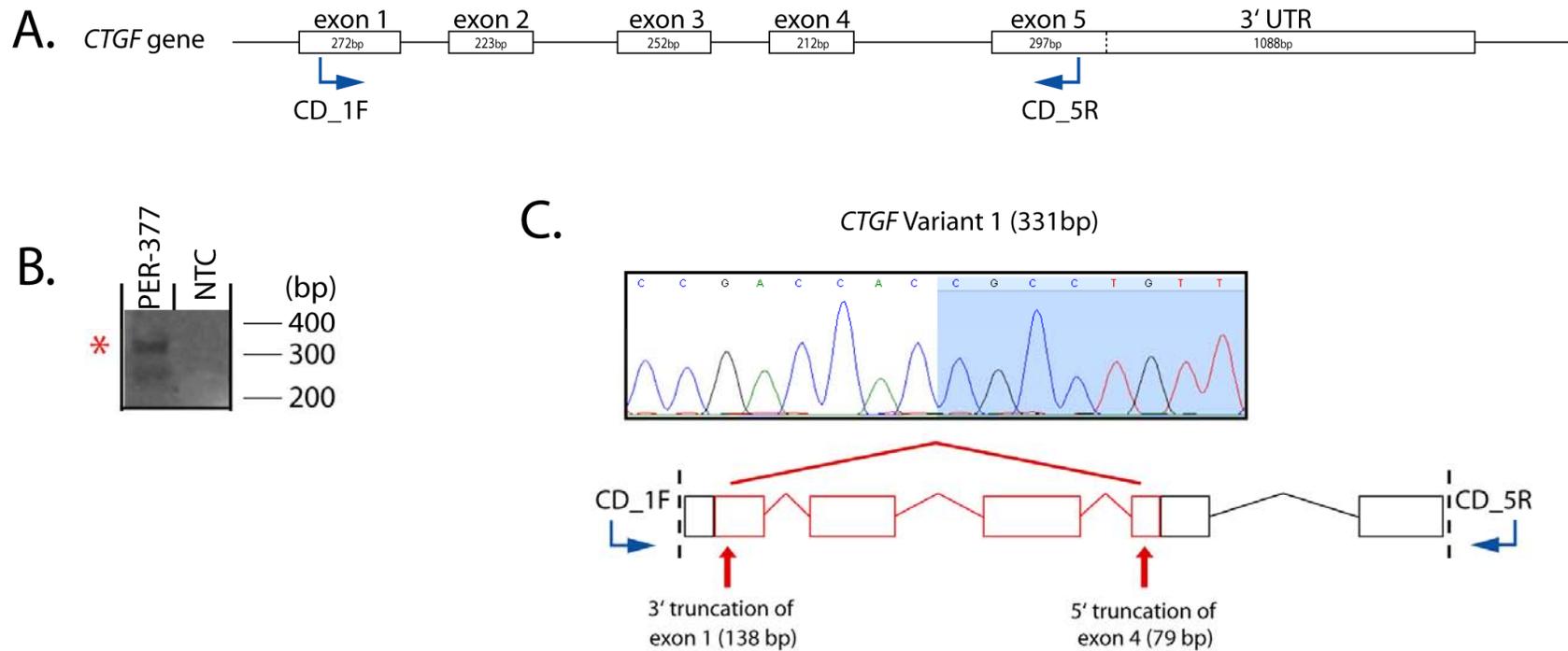


Figure 5.3 Structure of *CTGF* Variant 1

A. Schematic representation of the *CTGF* gene structure and location targeted by the CD_1F and CD_5R PCR primers. **B.** Amplification products generated from PER-377 cDNA using the CD_1F and CD_5R primers. The top product of 331bp indicated by a red asterisk was demonstrated to be a novel *CTGF* transcript (designated Variant 1) by sequencing. **C.** Sequencing the 331bp clone revealed skipping of exons 2 and 3 (shown in red) as well as truncation of exons 1 and 4.

Further extension of this PCR strategy was undertaken to capture other alternative splicing events affecting *CTGF* mRNA in PER-377 cells. A more sensitive semi-nested PCR approach was conducted by incorporating a primary round of amplification using a reverse primer within the 3' UTR of the *CTGF* transcript (CD_7R) shown in Figure 5.4A. This primary PCR was followed by a secondary, semi-nested PCR reaction using the CD_1F to CD_5R primer combination from the previous experiment. The results of this approach shown in Figure 5.4B, yielded a further two fragments of around 380bp and 400bp in addition to the previously identified Variant 1 (331bp). These larger PCR products were gel extracted and cloned, and sequencing these clones revealed that they too were *CTGF* transcript variants arising from alternative splicing events in addition to Variant 1.

These two new variants were found to be 366 bp (Variant 2) and 415 bp (Variant 3) respectively. Variant 2 exhibited splicing from a truncated exon 1, 35 bp downstream from the alternative splice site used in Variant 1. This shorter exon 1 was spliced to the same alternative acceptor splice site in exon 4 that was used to generate Variant 1, and splicing of exon 4 to exon 5 adhered to canonical exon junctions. The largest of these transcripts, Variant 3 was missing 170 bp of the 3' end of exon 1, and in contrast to Variants 1 and 2, this truncated exon 1 was spliced to a shortened form of exon 3. As with the other two variants, downstream splicing adhered to canonical exon junctions for this transcript. In all cases (Variants 1, 2 and 3) splicing of the alternative exon 1 donor splice site occurred prior to start of the canonical *CTGF* reading frame at position 207 of exon 1, thus the *CTGF* reading frame is 5' truncated in these alternative transcripts.

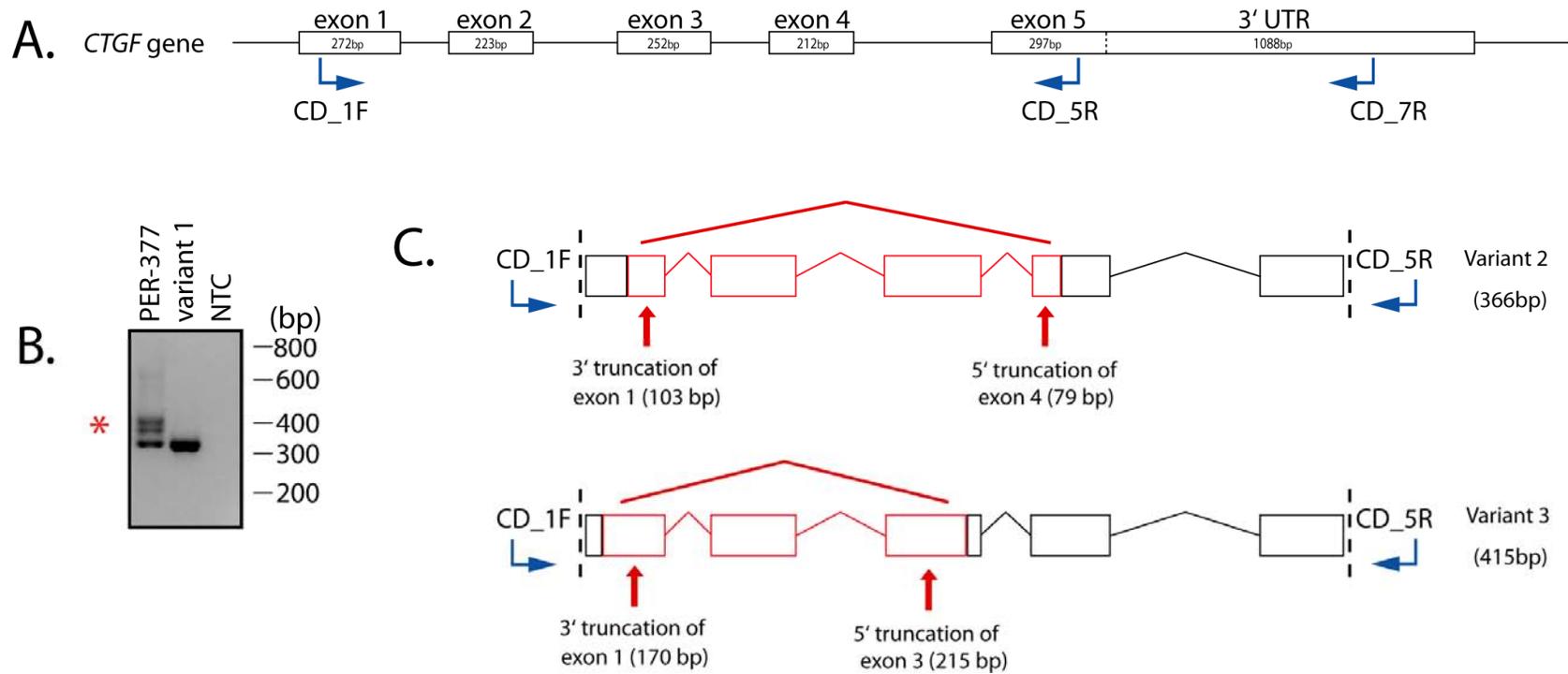


Figure 5.4 Structure of *CTGF* Variants 2 and 3

A. Schematic representation of *CTGF* gene structure and binding sites for the CD_1F, CD_5R and CD_7R PCR primers. **B.** Second round semi-nested amplification from PER-377 cDNA using CD_1F, CD_5R primers. Plasmid DNA containing the previously detected Variant 1 clone was used as a positive control in the secondary PCR. The top two bands denoted by a red asterisk were demonstrated to be *CTGF* transcript variants by sequencing. **C.** Schematic representation of alternative splicing of Variant 2 and Variant 3 *CTGF* cDNA.

5.2.2.1 Structural analysis of novel *CTGF* transcript variants

The structure of the three novel *CTGF* transcript variants outside of the region amplified by PCR was not known. However, if these three variant cDNAs retained the canonical transcriptional start and polyadenylation sites then their respective sizes would be; 1.67 kb (Variant 1), 1.7 kb (Variant 2) and 1.75 kb (Variant 3), thus any one, or all of these transcripts could represent the upper, non-canonical *CTGF* bands of approximately 1.6 kb and 1.8 kb detected in the northern blots shown in Figures 5.1 and 5.2. Splicing of all three of the variant transcripts occurred 5' of the canonical start codon. However, *in silico* analysis predicted that these transcripts retain the 3' of the canonical *CTGF* reading frame at position 102 of exon 4 (Met215). Figure 5.5 outlines the sequenced regions of Variants 1 to 3 and the position of Met215 is shown in exon 4. There is a partial match with a Kozac consensus sequence (De Angioletti et al. 2004) (gccrccAUGG) adjacent to Met215, as there is at the canonical *CTGF* start codon, thus the Met215 codon may still serve as a valid site for translation initiation.

If translated, these three transcripts could all encode a truncated *CTGF* protein containing half of the TSP1 domain and all of the CT domain. Immunoblotting experiments performed in the previous chapter with the C-terminal Rabbit-A antibody only detected the full length *CTGF* protein (38kDa) in enriched supernatants from PER-377, thus these transcripts do not appear to be translated into protein. However, this may be accounted for by technical reasons as the level of this protein isoform may be below the assay detection limit. Alternatively, the predicted shorter *CTGF* protein may not expose the epitope targeted by the Rabbit-A antibody.

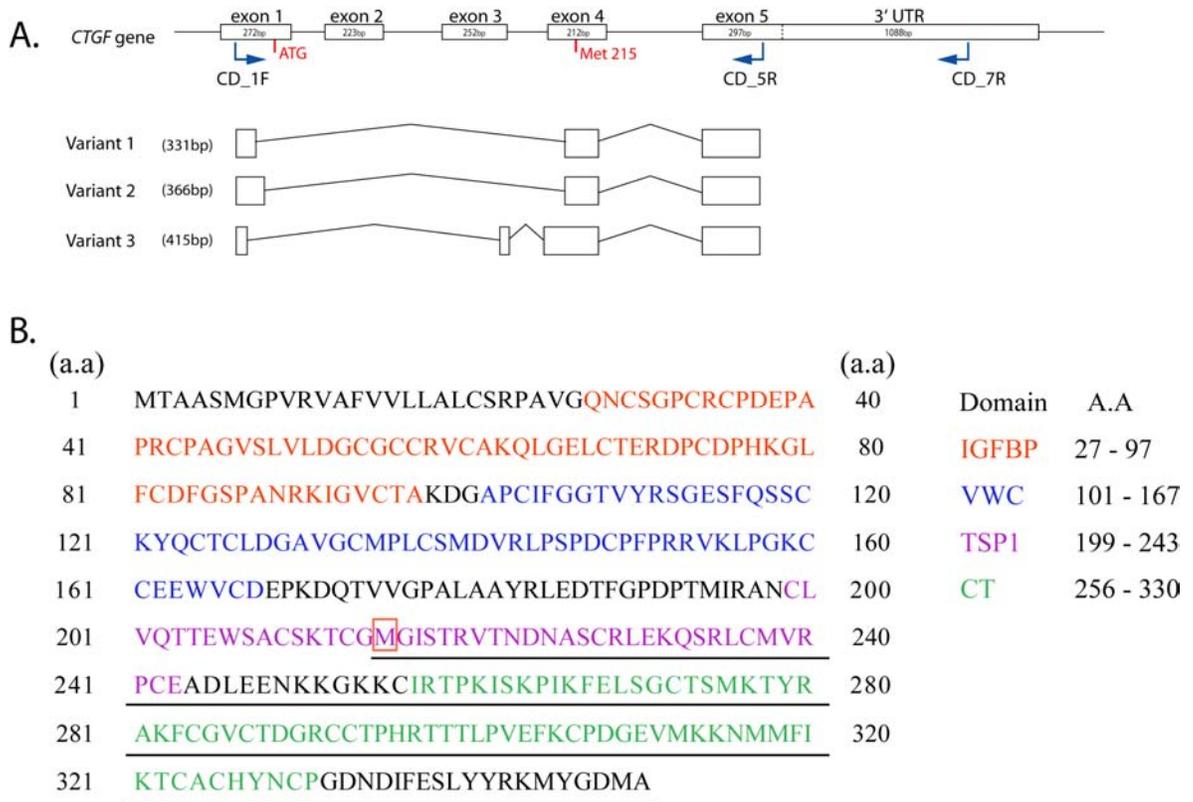


Figure 5.5 *CTGF* transcript variants may retain protein coding potential

A. *CTGF* gene structure with forward and reverse primers used to amplify Variants 1 to 3 indicated by blue arrows. A start codon (ATG) is encoded at position 207 of exon 1. Met215 encoded at position 102 of exon 4 is contained within all three variant cDNAs.

B. *CTGF* amino acid sequence. Functional domains are colour coded: IGFBP domain (red), VWC domain (blue), TSP1 domain (purple) and the CT domain (green). Translation initiation at Met215 (highlighted by the red box) can encode a 134 amino acid protein (underlined) corresponding to the C-terminal one-third of the *CTGF* protein.

5.2.3 Amplification of *CTGF* transcripts by RACE

RT-PCR amplification between exons 1 and 5 of *CTGF* from PER-377 RNA identified three novel cDNA fragments that were confirmed as *CTGF* splice variants by sequencing. It is possible that additional alternative transcripts are present in PER-377 other than those detected by the PCR approach. For this reason, 5' and 3' RACE (rapid amplification of cDNA ends) was performed in an attempt to detect the full repertoire of *CTGF* mRNA species and also to characterise the 5' and 3' ends of these variant transcripts. For both 5' and 3' RACE, two gene specific primers (GSPs) were used in nested RACE PCR reactions as outlined in Figure 5.6A. OligodT-primed first strand cDNA was generated from *DNaseI*-treated total RNA from PER-377 cells and served as a template for RACE PCR. Both 5' and 3' RACE PCR amplification from PER-377 cDNA yielded multiple RACE PCR products (Figure 5.6B). In the case of 5' RACE, the full length canonical transcript (1064bp) was not amplified in either primary or secondary PCRs, however several smaller products between ~700 bp and 200 bp were amplified. When 3' RACE was performed, there was strong amplification of a product at the expected size of the canonical *CTGF* transcript (2.4 kb) together with products at approximately 2.8 kb and 1 kb and less abundant products at 500 bp and 200 bp.

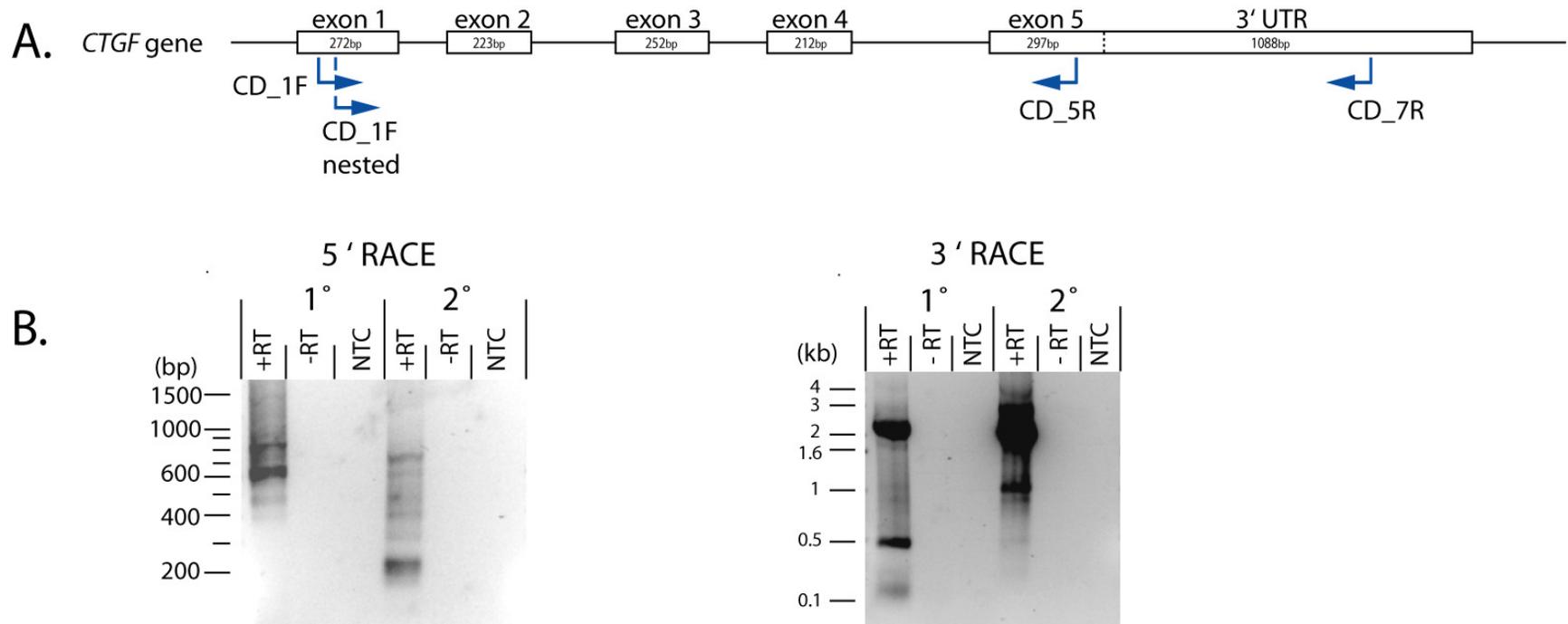


Figure 5.6 Amplification of *CTGF* variants by 5' and 3' RACE.

A. Location of gene specific primers (GSPs) used for RACE amplification. CD_1F and CD_1F_nested were used as GSPs for 3' RACE, while CD_7R and CD_5R were used as GSPs for 5' RACE. **B.** Amplification of PER-377 cDNA by 5' RACE and 3' RACE primary (1°) and secondary (2°) PCRs. Reverse transcriptase negative (-RT) controls from 1° PCR reactions were used as templates for 2° PCR (-RT) controls. No template controls (NTCs) were used for both reactions.

5.2.3.1 Isolation of RACE products by band-stab PCR

The distribution of RACE products shown in Figure 5.6B indicated that isolation of individual products would be difficult. In the case of 3' RACE, there was a clear amplification bias for the more abundant canonical transcript, which would result in an overrepresentation of this transcript if shotgun cloning was used to clone unselected RACE PCR products. In the case of 5' RACE, there were a large number of bands that did not resolve satisfactorily, even on high percentage agarose gels. For this reason, 5' and 3' RACE products of interest were isolated by using a band-stab PCR strategy (Bjourson & Cooper 1992). The band-stab PCR allowed the selective isolation of individual RACE products that would otherwise have been difficult to independently obtain as shown in Figure 5.7. For the 5' RACE PCR (Figure 5.7A), six bands were identified as independent PCR products and were targeted for re-amplification by band-stab PCR, using the GSP primer CD_5R and the secondary 5' RACE primer. For the 3' RACE PCR (Figure 5.7B), five bands were similarly selected and then re-amplified using the CD_1F_nested primer and the nested 3' RACE primer.

When 5' RACE products were re-amplified by band-stab PCR, the resulting products displayed the same size as those bands used as the template or target band in band-stab reactions (Figure 5.7A, lane 4, bands 1 to 6). Separation of these PCR products provided superior resolution compared to RACE PCR alone, facilitating the independent isolation of target bands by gel extraction. Separation of band-stab PCR products from 3' RACE revealed there was amplification of a product in lane 1 and 2 that corresponded to the size of the canonical *CTGF* transcript. Lane 3 contained a dominant product of approximately 1.5 kb, while lanes 4 and 5 contained dominant products of 600 and 250 bp respectively.

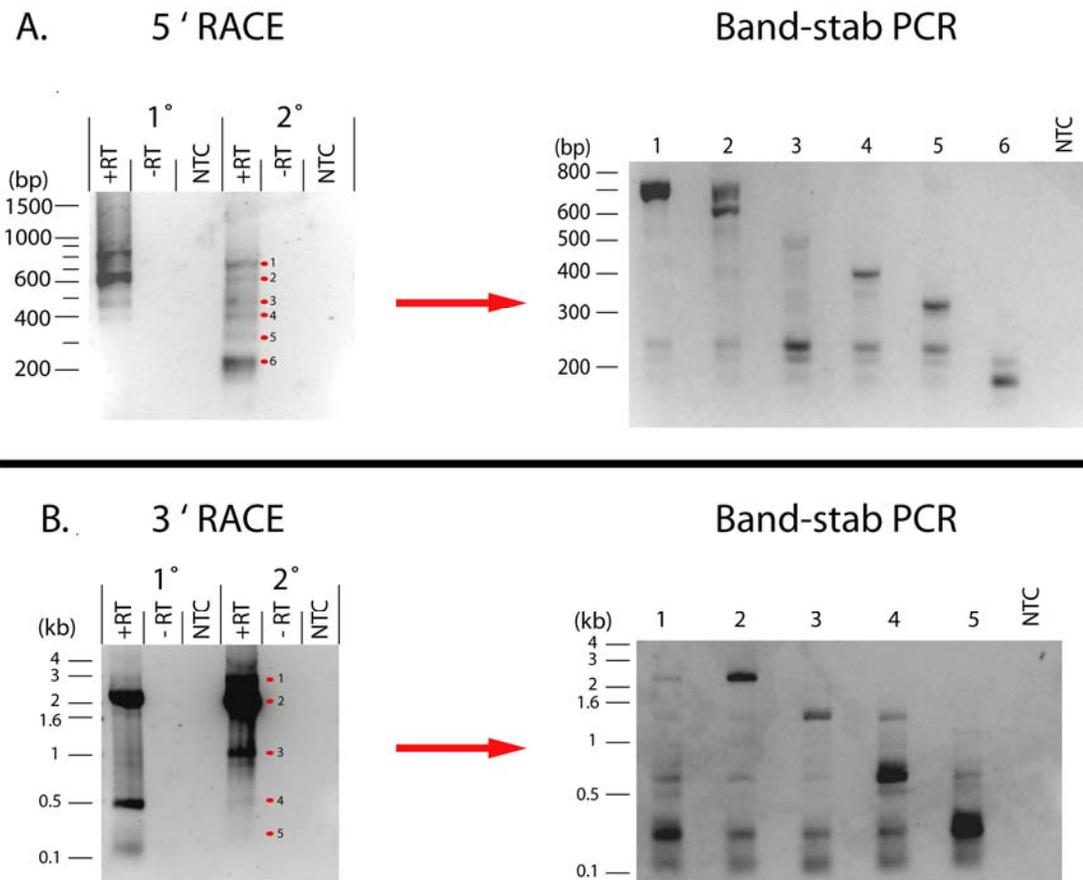


Figure 5.7 Re-amplification of individual *CTGF* RACE products by band-stab PCR

(A) 5' RACE products generated from PER-377 cDNA. Individual target bands 1 to 6 (red dots; left panel) from the nested secondary RACE PCR were re-amplified by band-stab PCR using the CD_5R GSP and the nested 5' RACE primer (right panel). (B) 3' RACE product generated from PER-377 cDNA. Individual target bands 1 to 5 (red dots; left panel) were re-amplified by band-stab PCR using the CD_1F_nested GSP and the nested 3' RACE primer (right panel).

5.2.3.2 Validation of a *CTGF* probe for Southern blot hybridisation

Before extracting and cloning RACE band-stab cDNAs, Southern hybridisation was used to verify that these products contained genuine *CTGF* sequences. Only those bands

hybridised by a radio-labelled *CTGF* specific probe were isolated for further analysis. The *CTGF* probe was generated by amplifying from exon 1 to 5 from a cloned, full length *CTGF* cDNA, using the CD_1F and CD_5R primers. The specificity of the *CTGF* probe was examined by hybridisation to unselected primary and secondary *CTGF* 5' RACE PCR products which displayed multiple bands by gel electrophoresis. As shown in Figure 5.8B, the *CTGF* probe strongly hybridised cDNA from the secondary RACE PCR only, indicating that the probe was highly specific for *CTGF* and was therefore suitable for probing band-stab PCR products.

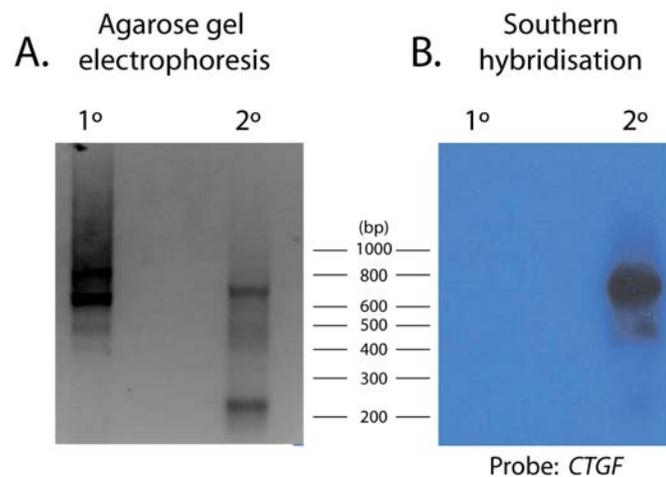


Figure 5.8 Southern blotting of 5' RACE PCR products

A. Primary (1°) and secondary (2°) PCR products from PER-377 5' RACE, separated by agarose gel electrophoresis and stained with ethidium bromide. **B.** Southern hybridisation of these samples, probed using a radio-labelled (³²P) *CTGF* probe.

5.2.3.3 Characterisation of 5' RACE products

Southern hybridisation of 5' RACE band-stab PCR products was performed with the *CTGF* probe to identify genuine *CTGF* RACE PCR products. As shown in Figure 5.9A, individual bands that were clearly hybridised by the *CTGF* probe (marked by red boxes) were deemed to be genuine *CTGF* RACE PCR products and were isolated and cloned.

Lane 2 in Figure 5.9A contains two prominent bands between 600 and 700 bp in size. The upper band was similar in size to the predominant product in lane 1 while the lower band did not appear to specifically hybridise the *CTGF* probe and as such, lane 2 products were not included in further analysis. A total of seven bands were isolated and cloned and in the case of lanes 3 and 6, two individual bands were isolated from each lane. Three independent clones from each transformation were isolated and sequenced. Sequencing of these twenty one 5' RACE clones revealed a total of eleven distinct *CTGF* transcripts, as shown in Figure 5.9B. Overall, there was a high degree of heterogeneity observed in the 5' end of 5' RACE clones and this heterogeneity suggests that there are likely to be additional *CTGF* transcripts present that were not fully characterised in these experiments. Of note, splicing of downstream intron/exon junctions were in agreement with canonical mRNA exon boundaries. Those 5' ends within exonic regions are colour coded in red, while transcripts displaying 5' ends localised within introns were similarly colour coded, however these were grouped according to their intron of origin, yellow for intron 2, blue for intron 3, and green for intron 4.

Sequencing these RACE products revealed that clones 5_1:c1-c3 were between 688 and 706 bp in size and exhibited 5' ends ranging from 89 to 108 bp upstream from the exon 3 boundary. Clones 5_3A:c1-c3 were isolated from the upper band in lane 3 of Figure 5.9A. All three clones were identical in size (476 bp) and structure, exhibiting a 5' end contained within intron 3, 129 bp upstream from the canonical exon 4 boundary and only 3bp downstream from the canonical exon 3 donor splice site. The 5_3B:c1-c3 clones derived from isolates of the lower band in lane 3 of Figure 5.9A were also identical in size and structure. These cDNAs were 222 bp in length and exhibited a 5' start site 89 bp upstream from the canonical end of exon 4.

Clones derived from lane 4 in Figure 5.9A were similar in size but differed in their 5' ends and structure. Clones 1 and 2 (5_4:c1,c2) both exhibited 5' ends originating within exon 3, just before the canonical exon 3 donor site boundary, while the 5' end of clone 3 (5_4:c3) originated from within intron 3, 48 bp upstream from the canonical exon 4

acceptor site. The three 5_5:c1-c3 clones were 326 bp in size and exhibited a 5' end originating within intron 4, 191 bp upstream from the 5' end of exon 5. The upper isolate from lane 6 in Figure 5.9A (55_6A) was 197 bp in length and contained 62 bp of the 3' end of exon 4 spliced to exon 5. All three clones were identical. Clones derived from 5_6B were also identical and contained 36 bp of exon 4 spliced to exon 5. None of these 5' RACE clones contained any part of exons 1 or 2 in their 5' end. This may be for technical reasons since, as stated previously, the canonical *CTGF* cDNA could not be obtained by 5' RACE and exons 1 and 2 have a high GC content. Similarly, the cDNA variants detected by RT-PCR in the previous section were not detected by 5' RACE.

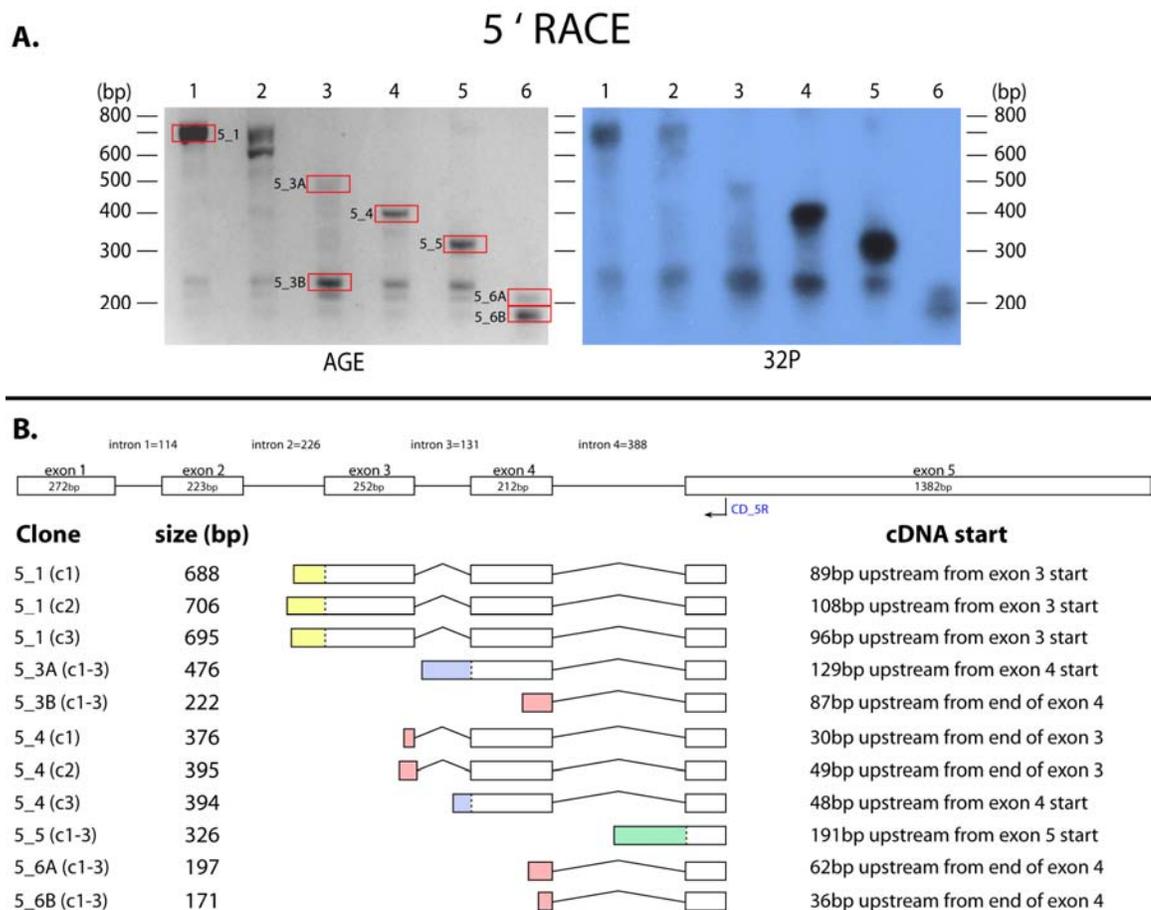


Figure 5.9 Characterisation of PER-377 *CTGF* 5' RACE products

A. Band-stab isolated RACE products separated by agarose gel electrophoresis (AGE) and then examined by Southern blotting with a radio-labelled (^{32}P) *CTGF* probe. Red boxes indicate bands that were isolated, cloned and sequenced. Where two or more isolates were obtained from a single lane, isolates are labelled by adding A or B to name.

B. Schematic representation of 5' RACE clones aligned with the *CTGF* gene. Reverse primer CD_5R is shown in blue. Clone name is followed by identifier for sequenced clone (c1, 2 or 3), and the size of clones is shown in bp. The cDNA start site is indicated on the right of each RACE clone. Vertical broken lines in transcripts represent canonical exon junctions. Clones with 5' ends originating within canonical *CTGF* exons have their first exon coloured in pink, while those displaying 5' ends within intronic regions have the intronic portion of their 5' end colour coded as yellow for intron 2, blue for intron 3 and green for intron 4.

5.2.3.4 Characterisation of 3' RACE products

Southern hybridisation of 3' RACE PCR products re-amplified by band-stab PCR (Figure 5.10A) was performed as previously described for 5' RACE PCR products. Hybridisation with the *CTGF* specific probe revealed that lane 1 did not contain any *CTGF* positive products that were unique compared with other lanes and as such, this lane was excluded from further analysis. A single band was extracted from each of the remaining 4 band-stab PCR reactions (lanes 2 to 5) as indicated by the red boxes in Figure 5.10A. Sequencing of the cloned isolates was performed with the primers CD_4F, CD_5F, CD_6R and CD_7R which anneal within exonic regions of the *CTGF* gene as indicated in Figure 5.10A. The 3' end of cloned band-stab isolates were sequenced using vector specific primers (M13 forward and reverse).

Sequencing of the 2.3 kb clones derived from lane 2 in Figure 5.10A (3_2:c1-c3) revealed that this 3' RACE PCR product was amplified from the full length, canonical *CTGF* transcript. This is in contrast to 5' RACE which did not yield products amplified from the canonical *CTGF* transcript. Analysis of 3_3:c1-c3 clones revealed that this isolate was amplified from a novel *CTGF* transcript that exhibited premature polyadenylation within exon 5. This transcript contained 384 bp of exon 5 and harboured an alternative polyadenylation site located 87 bp downstream from the canonical translation termination codon. Thus, this transcript retained *CTGF* coding potential, but was devoid of established regulatory elements contained within the 3' UTR (Ohgawara et al. 2009; Duisters et al. 2009; Ernst et al. 2010; Kubota et al. 2005). If transcribed from the canonical transcription start site, this mRNA variant would be 1.36 kb in length, thus this transcript may represent the lower band detected by northern blotting of B-lineage ALL cell lines, see Figures 5.1 and 5.2. This mRNA variant may represent a long-lived transcript with increased translational potential due to the lack of the 3' regulatory elements. Thus, while this minor transcript was only present at low concentrations, it may nevertheless have biological importance.

Sequencing of 3_4:c1-c3 demonstrated that these isolates were cloned from a PCR artefact generated through reverse priming by the CD_1F nested primer, and not a genuine RACE product. Reverse priming occurred despite low specificity between primer and target, with only five bases of complementarity wholly localised to the 3' end of the primer. Sequencing of the final RACE PCR isolate (3_5:c1-c3) revealed a novel RACE product that contained 121bp of *CTGF* exon 1, followed by a 46 bp of unknown sequence and a putative polyadenylation signal (shown in more detail in Figure 5.11A). Using the BLAST sequence alignment tool, this unknown sequence was used to query the human RNA_refseq database for a potential sequence match. Only one mRNA transcript; leucine rich repeat containing 58 (*LRRC58*) mRNA, matched the sequence query with 100 percent identity, shown in Figure 5.11B. Furthermore, when the 3_5 clone sequences were aligned with the *LRRC58* genomic sequence (Figure 5.11C), it was apparent that the *LRRC58* genomic sequence contained a run of adenines similar to those seen in the sequenced RACE clone indicating that this segment of the sequenced clones may not represent the 3' polyadenylation site of the original mRNA, thus the full length mRNA transcript may have been considerably longer.

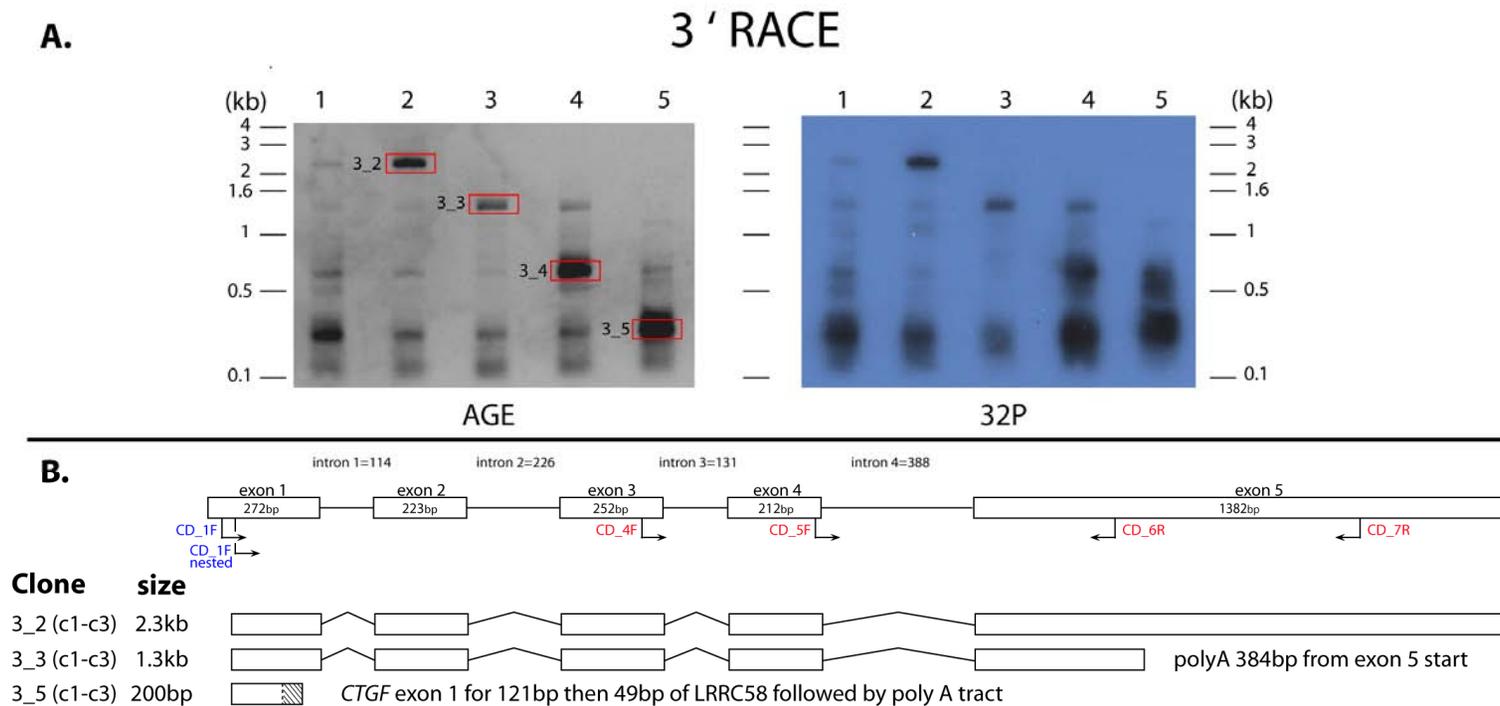


Figure 5.10 Characterisation of PER-377 *CTGF* 3' RACE products

A. Band-stab isolated RACE products were separated by agarose gel electrophoresis (AGE) and then examined by Southern blotting with a radio-labelled (^{32}P) *CTGF* probe. Red boxes indicate bands that were isolated, cloned and sequenced. **B.** Schematic representation of 3' RACE clones aligned with the *CTGF* gene. Clone ID and size are indicated on the left. Forward GSPs for RACE PCR shown in blue. Primers used for internal sequencing of clones shown in red. M13 forward and reverse vector primers were used to sequence insert ends.

These observations suggested that there may be a translocation or insertion event involving the *LRR58* and *CTGF* gene loci. This seemed unlikely given the results reported Chapter 4 that ruled out rearrangement of the *CTGF* locus in PER-377, however to test this possibility two reverse primers targeting the *LRR58* locus immediately upstream from the run of adenines in Figure 5.11C were used in conjunction with the CD_1F and CD_1F_nested PCR primers. It was hypothesised that if the 3_5 RACE clone arose from a fusion between these two genomic loci these primers would facilitate amplification of the fusion gene. Amplification with these primer combinations using PER-377 genomic DNA did not yield any products, ruling out the existence of such a fusion in this cell line. Therefore this fusion transcript most likely arose as a result of aberrant trans-splicing between two disparate pre-mRNA transcripts and is unlikely to represent a genuine coding transcript (Houseley & Tollervey 2009; Li et al. 2009b).

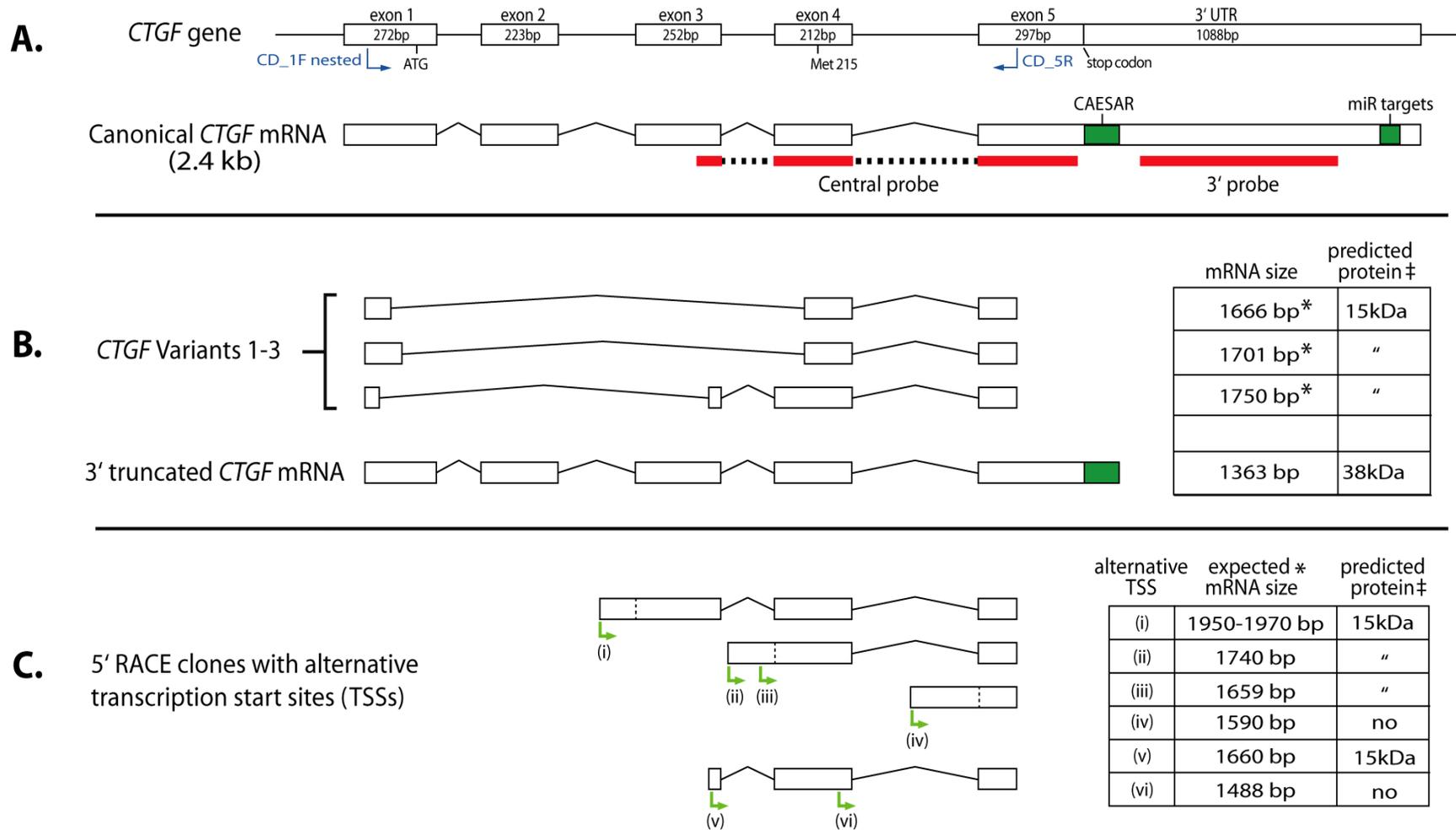


Figure 5.12 Summary of novel *CTGF* transcripts detected in pre-B ALL cells (Figure legend over page)

Figure 5.12 Summary of novel *CTGF* transcripts detected in pre-B ALL cells

A. Schematic representation of canonical *CTGF* mRNA aligned with the *CTGF* gene. Forward and reverse primers used in RACE experiments are shown in blue. The Met 215 codon which may serve as a secondary translation start site, is shown in exon 4. The Central and 3' northern blot probes (red bars) are aligned to the *CTGF* mRNA transcript. Post-transcriptional regulatory elements contained within the *CTGF* 3' UTR are shown in green. **B.** Novel *CTGF* transcripts detected by RT-PCR (Variants 1 to 3) are shown together with the 3' truncated RACE transcript lacking most of the 3' UTR. For each of these transcripts, the mRNA size is shown. *Predicted assuming canonical TSS and/or polyadenylation sites were present outside of sequenced regions. The predicted protein coding potential of transcripts is also shown ‡*In silico* predictions of protein coding potential assumed the presence of an intact exon 5 including canonical stop codon. **C.** Representation of 5' RACE clones identified in this study that exhibited non-canonical 5' ends. Vertical broken lines represent canonical exon boundaries. Transcripts are grouped by the location of their 5' ends, either intron2, intron 3, intron 4 or those transcripts that originated within exonic regions. For each of the four groups, alternative transcription start sites (alt. TSS) are shown by numbered green arrows (i to vi) and annotated on the right, with expected mRNA size* and predicted protein coding potential‡.

5.3 Discussion

The disparity between *in vitro* *CTGF* mRNA and protein levels documented in Chapter 4 prompted a more detailed analysis of *CTGF* mRNA to confirm that those *CTGF* transcripts detected by qRT-PCR and microarray methods represented the canonical 2.4 kb message. The present findings demonstrate that the canonical *CTGF* mRNA was indeed the major form of *CTGF* transcribed in B-lineage ALL cell lines. Furthermore, canonical *CTGF* transcripts were readily detectable in the *CTGF*^{low} cell lines PER-145, PER-278 and PER-371 by northern blotting. This confirmed that these *CTGF*^{low} cell lines generate *CTGF* mRNA, despite lacking detectable levels of CTGF protein, as documented in Chapter 4.

In addition to canonical *CTGF* mRNA, northern blotting revealed the presence of shorter, much less abundant *CTGF* transcripts produced in *CTGF* positive cell lines. Sequencing of these mRNA variants (isolated from the *CTGF*^{high} cell line PER-377), confirmed that these minor *CTGF* transcripts, estimated to be present at less than 10% of the total pool of *CTGF* mRNA, were produced by alternative pre-mRNA splicing. This is the first evidence of alternative splicing of *CTGF* mRNA in any cell type and supports further investigation into alternative *CTGF* splicing in other cancers and normal tissues. A number of alternative splicing events were observed in sequenced cDNA fragments and RACE PCR products generated from the PER-377 cell line. These non-canonical transcripts summarised in Figure 5.12, exhibited exon skipping, alternative 5' (donor) and 3' (acceptor) splice site usage, alternative or premature polyadenylation and alternative transcription start sites.

5.3.1 Aberrant *CTGF* transcription

Targeted RT-PCR amplification of *CTGF* cDNA from exon 1 to 5 revealed the presence of three novel transcripts (Figure 5.12B). These transcripts exhibited internal deletions of exons 2 and 3, alternative exon 1 donor splice sites, and alternative exon 3 and 4 acceptor splice sites. *In silico* analysis of these alternatively spliced transcripts indicated that they have the potential to encode a shorter, N-terminal truncated *CTGF* protein of approximately 15kDa. This protein would lack N-terminal domains including the signal peptide which may impair secretion of this truncated *CTGF* protein. Several smaller MW *CTGF* proteins ranging in size from 10 kDa up to 20 kDa have been detected in human cells and biological fluids (Steffen et al. 1998; Yang et al. 1998), however their existence has uniformly been attributed to post translational proteolysis of the full length *CTGF* protein. This notion may need to be reconsidered given the current findings, particularly if alternative splicing of *CTGF* is demonstrated in other cell types. The ability of these truncated *CTGF* cDNAs to encode biologically active proteins should be examined *in vitro*. If such protein isoforms are confirmed, functional characterisation will be required to establish their biological function and sub-cellular localisation.

In addition to alternative splicing events identified by RT-PCR, analysis of *CTGF* transcripts by 3' RACE uncovered a novel 1.3 kb *CTGF* variant with an early polyadenylation signal 87 bp after the end of the *CTGF* coding region (Figure 5.12B). This transcript contained an intact coding region, but did not contain the canonical 3' UTR and thus was devoid of important regulatory elements that have been shown to contribute to the post-transcriptional regulation of *CTGF* expression in other cell types. These include target sites for miRNAs, including miR18a, miR30, miR130 and the miR-17-92 cluster (Ohgawara et al. 2009; Duisters et al. 2009; Ernst et al. 2010). While these miRNA target sites were missing from this 1.3 kb transcript, the CAESAR element described by Kubota and colleagues remained intact (Kubota et al. 2005). If this transcript was initiated at the canonical TSS, which is likely given the proximity of the gene specific forward primer CD_1F to the annotated *CTGF* TSS (Figure 5.12A), then the predicted size of this transcript would be 1.362 kb. This is the approximate size of one of the non-canonical transcripts identified by northern blots and points to the likely abundance of this transcript in *CTGF* positive B-lineage ALL cell lines, albeit at a minor frequency. Increased early polyadenylation of mRNA transcripts is a general feature of cancer cells and has been shown to enhance the protein coding potential of transcripts when compared to canonical mRNA by up to ten-fold (Mayr & Bartel 2009). Recombinant expression of this cDNA may reveal a higher translational affinity compared to canonical *CTGF* owing to differences in the 3'UTR. Thus, this 1.3 kb *CTGF* transcript represents an important candidate for further investigation and efforts should be made to determine its biological distribution in ALL cells and other tissues.

Analysis of *CTGF* mRNA conducted by 5' RACE detected a range of cDNAs with different 5' ends ranging from internal regions of intron 2 to intron 4, summarised in Figure 5.12C. These non-canonical 5' ends may represent alternative TSSs, suggesting possible promoter activity within the *CTGF* coding region. Analysis of 5' RACE clones indicated that there were six regions that served as alternative TSSs (shown as green arrows in Figure 5.12C), one in intron 2, two in intron 3 one in intron 4 and on each in the 3' end of exons 3 and 4. As with the *CTGF* cDNA fragments identified by RT-PCR,

some of these novel transcripts are predicted to encode shorter CTGF protein isoforms if translation is initiated at Met 215 and the coding portion of exon 5 remains intact. A limitation of the present study is that 5' RACE clones were amplified using a GSP upstream from the canonical stop codon (CD_5R), and thus the downstream composition and coding potential of the 5' RACE clones is not certain.

A recent report suggests that the vast majority of protein coding genes are affected by alternative TSS usage, however the causes behind this phenomenon remain unclear (Denoëud et al. 2007). Alternative TSS selection has not been reported to affect human *CTGF* transcription, however a recent report by Huang and colleagues demonstrated that the murine *CTGF* 5' UTR can induce reporter gene activation through promoter-like sequence elements (Huang et al. 2007). While the 5' end of the human *CTGF* gene has not been investigated for promoter-like activity, the present findings suggest that such a study may be warranted as this may account for the alternative TSSs observed in the present study.

5.3.2 Post transcriptional regulation of *CTGF* mRNA

In assessing biological specimens, and in particular large patient cohorts for expression of a gene or genes of interest, microarray technology has paved the way for large-scale, high-throughput analysis. However, despite the importance of gene expression data, recent studies have suggested that mRNA levels are only a weak surrogate for the expression of proteins they encode (Tian et al. 2004; Nie et al. 2006), capturing only around 40% of global changes in protein expression levels. In agreement with this notion, *CTGF* mRNA was not a direct predictor of whether ALL cells actively synthesise CTGF protein. This may have been for technical reasons, because while low mRNA levels can be detected by RT-PCR and northern blot, correspondingly low protein levels may be below the detection limit of immunoblotting. However, another distinct possibility is that post-transcriptional regulation impacts upon the active synthesis of CTGF protein in these *CTGF*^{low} cell lines. Post-transcriptional regulation of gene expression has emerged as of key importance in recent years (Volinia et al. 2010;

Ruan et al. 2009), and there is significant evidence that miRNAs can modulate CTGF protein expression (Ohgawara et al. 2009; Duisters et al. 2009; Ernst et al. 2010). Investigation of miRNA levels in ALL was considered beyond the scope of the present study. However, analysis of the levels and activity of *CTGF* associated miRNAs may reveal an association between their activity and CTGF protein levels in pre-B ALL and such a study should be considered in the future.

The importance of being able to accurately classify the CTGF protein status of ALL cells is highlighted by a 2007 study conducted by Sala-Torra and colleagues (Sala-Torra et al. 2007). Analysis of adult ALL specimens revealed that a poor outcome was associated with high *CTGF* expression. Of particular note in light of the current findings however, is that when patients were grouped into low, intermediate or high *CTGF* mRNA expression groups, the 5 year overall survival of patients was 58, 42 and 11% respectively (Sala-Torra et al. 2007). These data indicate that patients with high levels of *CTGF* mRNA exhibited markedly poorer survival compared to those with low or intermediate levels of *CTGF* mRNA. The average survival rate for adults diagnosed with ALL is around 40% (Rowe et al. 2005; Annino et al. 2002), thus the group with high *CTGF* mRNA had a significantly poorer outcome.

The Sala-Torra report, coupled with the findings made in the present study suggest that there may be a stochastic mRNA tipping point that must be achieved before CTGF protein synthesis can occur. It is feasible that as the level of *CTGF* mRNA increases, the ability of post-transcriptional regulatory elements to hold CTGF protein synthesis in check is curtailed. No analysis of CTGF protein expression in pre-B ALL patient specimens has been reported to date and it is clear that such a study is now warranted to correlate *CTGF* mRNA with protein levels and furthermore to ascertain if stratifying patients based upon CTGF protein expression holds more robust prognostic significance compared to mRNA levels alone.

5.3.3 Conclusions

The experiments reported in this chapter confirmed that while the major form (>90%) of *CTGF* mRNA expressed in B-lineage ALL cell lines was canonical, there was also a range of alternative *CTGF* transcripts produced in these cell lines including one transcript missing 3' regulatory elements. Furthermore, the observation that *CTGF*^{low} cell lines transcribe canonical *CTGF* suggests that post-transcriptional regulation may be limiting CTGF protein synthesis in these cell lines. These findings are the first sequence-level evidence of alternative splicing of *CTGF* mRNA and should prompt future research aimed at elucidating the full extent of alternative splicing affecting *CTGF* expression in both healthy and diseased tissues.

Chapter 6

Epigenetic Mechanisms

Influencing *CTGF* Expression in

Pre-B ALL

Chapter 6

Epigenetic Mechanisms Influencing *CTGF* Expression in Pre-B ALL

6.1 Introduction

Epigenetic changes are now accepted as a hallmark of many cancers, including those of hematopoietic origins (Bonifer & Bowen 2010; Martin-Subero et al. 2009). Cytosine methylation represents a reversible, epigenetic change that occurs predominantly at CpG dinucleotides in the mammalian genome, and was first suggested to play a role in gene expression and differentiation in the mid 1970's (Holliday & Pugh 1975; Riggs 1975). The notion that methylation of cytosine residues can result in altered gene expression without changing the underlying DNA sequence is now widely accepted, but still poorly understood (Watanabe & Maekawa 2010). This is evidenced by the recent discovery of 5'-hydroxy-methylcytosine (OH-m⁵C) which is indistinguishable from 5'-methylcytosine (m⁵C) by traditional bisulfite conversion-based analysis and is postulated to have functions distinct from that of m⁵C (Kriaucionis & Heintz 2009; Nestor et al. 2010). As well as covalent modifications to DNA, post-translational modification of histones, the fundamental building blocks of the nucleosome, can alter their interaction with DNA and other nuclear proteins, impacting significantly upon the transcriptional potential of surrounding DNA. A large number of histone modifications with the ability to alter gene expression have been characterised including acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ADP ribosylation, deimination and proline isomerisation (Kouzarides 2007). Together, DNA methylation and histone modifications represent an epigenetic code with enormous biological importance.

Genomic hypomethylation was the initial epigenetic abnormality identified in human cancers almost 30 years ago (Feinberg & Vogelstein 1983a; Gama-Sosa et al. 1983). These seminal reports identified a significant decrease in the overall m⁵C content in primary cancer cells compared to healthy cells isolated from adjacent tissue. Furthermore, this reduction in m⁵C was progressive and further enriched in metastases

compared with the original tumours, suggesting that increased hypomethylation was associated with a more aggressive or metastatic phenotype (Feinberg & Vogelstein 1983a; Gama-Sosa et al. 1983). A raft of subsequent studies have demonstrated genomic hypomethylation in numerous cancer types, including prostate, bladder, cervical, and brain, as well as chronic lymphoblastic leukaemia (Brothman et al. 2005; Seifert et al. 2007; de Capoa et al. 2003; Cadieux et al. 2006; Wahlfors et al. 1992). A global reduction in DNA methylation is likely to significantly impact upon gene expression within cancer cells by promoting transcriptional activation of regions that are normally silent.

Aberrant chromatin changes are frequently observed in MLL-rearranged leukaemia and contribute to leukaemogenesis through epigenetic activation of *HOXA* and *MEIS1* genes, which promote uncontrolled self-renewal (Guenther et al. 2008; Somerville & Cleary 2010). The role of DNA methylation in leukaemogenesis is less clear, however it is generally accepted that DNA methylation can inhibit binding by trans-factors to DNA and can also recruit protein complexes that remodel chromatin and promote gene inactivation (Feng & Zhang 2001). Thus, an increase in DNA methylation at gene promoters is generally associated with inactivation of gene transcription and may contribute to leukaemogenesis by silencing of tumour suppressor loci (Vucic et al. 2008; Bernstein et al. 2007).

Studies reporting hypermethylation of classical tumour suppressor genes (TSG) have dominated the scientific literature (Zhang et al. 2007; Sidhu et al. 2005; Ebinger et al. 2004; Honorio et al. 2003), however almost all cases of TSG silencing through focal hypermethylation is accompanied by a general hypomethylation of the cancer genome, and these changes can result in activation of proto-oncogenes (Ehrlich 2009). Feinberg and colleagues provided the first evidence of this when they identified hypomethylation of *HRAS* and *KRAS* in lung and colonic adenocarcinomas compared to normal tissue (Feinberg & Vogelstein 1983b). Activation of proto-oncogenes associated with a reduction in DNA methylation in gene regulatory bodies has now been recorded in numerous cancer types, including prostate, liver, chronic lymphocytic leukaemia and

carcinoma (Wang et al. 2007; Tsujiuchi et al. 1999; Hanada et al. 1993; Smith et al. 2009).

Experiments reported in the previous chapters did not identify mutations or genetic changes that could account for the aberrant *CTGF* expression in pre-B ALL. Therefore, it may be possible that changes in *CTGF* gene expression arise due to altered CpG methylation at the *CTGF* locus. As discussed in Chapter 1, there is some evidence for the involvement of epigenetic phenomena in the regulation of *CTGF*. An inverse correlation between DNA methylation at the *CTGF* locus and *CTGF* gene expression has been demonstrated in ovarian and liver cancers (Kikuchi et al. 2007; Chiba et al. 2005). Furthermore, a recent study investigating the methylome in a large panel (n=367) of haematological neoplasms, identified hypomethylation at the *CTGF* locus, in 42 pre-B ALL specimens and hypermethylation of this locus in 18 T-ALLs, as well as 54 diffuse large B-cell lymphomas (DLBCL) (Martin-Subero et al. 2009). This is compelling evidence for the involvement of epigenetic phenomena in deregulation of the *CTGF* locus, however the high throughput array used in this study only targeted a small number of CpGs (1505 individual CpG sites in 807 genes), and only 2 CpGs were interrogated at the *CTGF* locus. The first CpG designated as P693, was located within the AP-1 binding site of the *CTGF* promoter at position -638, and was de-methylated in pre-B ALL specimens. The second site designated as E156 was located within exon 1 at position +212 and was methylated in T-ALL and DLBCL specimens (Martin-Subero et al. 2009). The authors of this study did not investigate whether *CTGF* gene expression correlated with the differential methylation of these CpG residues. Collectively, these studies point to a significant role for epigenetic regulation of *CTGF* expression. Thus, further more detailed investigation into the epigenetic state of the *CTGF* locus in leukaemia is warranted.

A comprehensive assessment of the methylation status of the *CTGF* locus was performed in cell lines and primary specimens to examine if aberrant CpG methylation may contribute to deregulation of the *CTGF* locus in pre-B ALL. The specific aims of the experiments presented in this chapter were as follows: 1) to investigate whether the

methylation status of the *CTGF* locus correlated with *CTGF* mRNA levels in cell lines, 2) to investigate if such a correlation existed in primary patient specimens, 3) and finally to examine whether modulating the global epigenetic state of cells *in vitro* could influence *CTGF* gene expression.

6.2 Results

6.2.1 Identification of a CpG island at the human *CTGF* genomic locus

To evaluate the distribution of CpG dinucleotides at the *CTGF* locus, the promoter and coding region from -800 to +3200 (relative to the TSS) was scrutinized using the “CpG island Searcher” program <http://www.uscnorris.com/cpgislands2/cpg.aspx> (Takai & Jones 2003). This program can detect an overrepresentation of CpG dinucleotides (CpGs) in DNA sequences defined by the user. The analysis identified a CpG island of approximately 1400 bp spanning the *CTGF* proximal promoter and extending to within exon 3 (Figure 6.1). Closer inspection revealed that the region containing the highest density of CpGs spanned position -91 to +883, comprising 111 individual CpG residues beginning around 100bp upstream from the canonical TSS and ending at the 5' end of exon 3. The identified *CTGF* CpG island is shown in more detail in Figure 6.2.

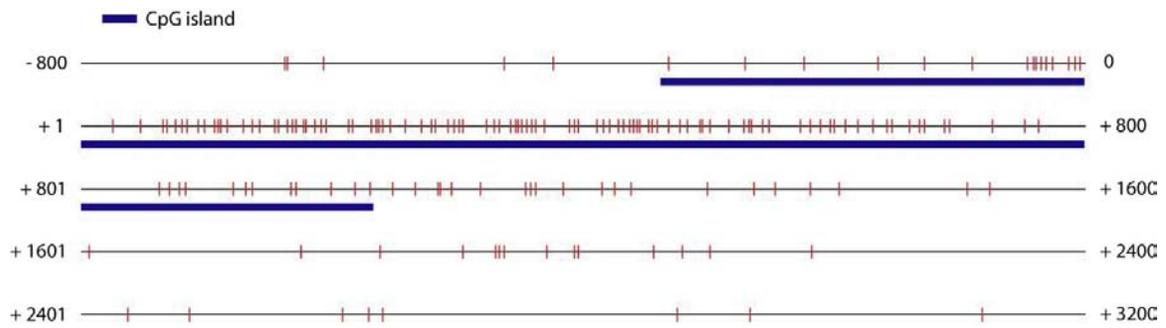


Figure 6.1 CpG dinucleotide distribution within the *CTGF* locus

CTGF genomic DNA sequence from -800 to +3200 was examined for distribution of CpG dinucleotides using the online tool “CpG Island Searcher”(Takai & Jones 2003). Nucleotide positions are indicated relative to the transcription start site (TSS). Red vertical lines denote identified CpGs. The blue bar indicates the extent of a CpG island within the *CTGF* locus.

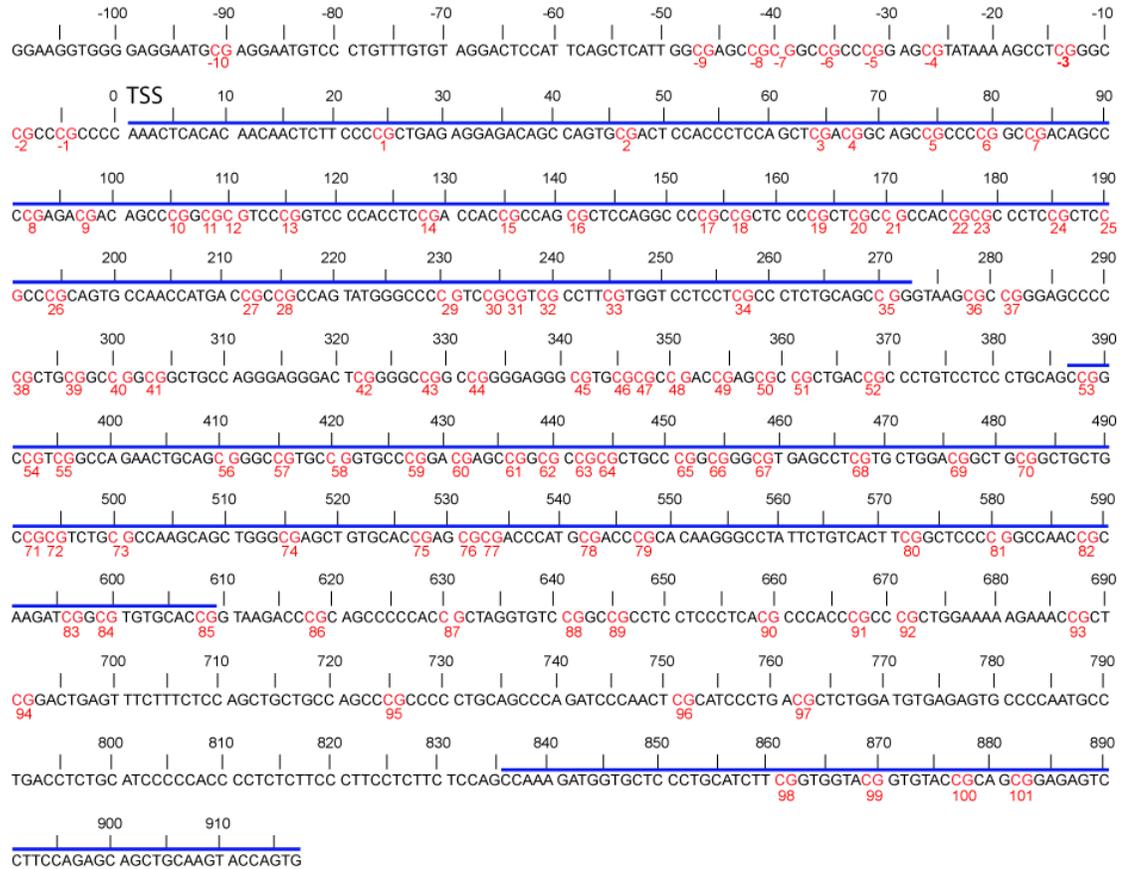


Figure 6.2 Nucleotide sequence of the *CTGF* CpG island

The *CTGF* CpG island extends from position -91 to +883 of the *CTGF* genomic locus. Blue bars denote exons 1 to 3. CpGs are shown in red and are numbered based upon position relative to the TSS.

6.2.2 Methylation correlates with *CTGF* expression in ALL cell lines

To determine if there was a correlation between CpG methylation and *CTGF* gene expression, methylation-specific PCR (msPCR) was performed on bisulfite-treated DNA from eight ALL cell lines, whose *CTGF* status had been established previously by qRT-PCR (see Chapter 4). This panel of cell lines included seven B-lineage ALL cell lines, four of which were positive for *CTGF* mRNA, as well as the T-ALL cell line (JURKAT) that was used as a negative control due to the absence of *CTGF* expression in T-ALLs (Boag et al. 2007). The primers used in the msPCR assays were originally described by

Chiba and colleagues to investigate the methylation status of the *CTGF* locus in hepatoma specimens (Chiba et al. 2005). Figure 6.3A outlines the design of the msPCR reactions. DNA was treated with bisulfite which converts cytosines to uracil via a deamination reaction. Methyl-cytosines were protected from conversion and site-specific primers allowed discrimination between methylated (protected) or unmethylated (converted) DNA using the methylated-site PCR (MSP) and unmethylated-site PCR (USP) primer pairs. These two primer pairs annealed to and allowed amplification from CpGs 5 to 8 at the forward priming site and 27 to 29 at the reverse priming site as outlined in Figure 6.3A. based upon their methylation status, generating a 159 bp PCR product.

The expression level of *CTGF* mRNA in the ALL cell lines is shown in Figure 6.3B. All four of the *CTGF* positive cell lines PER-145, PER-278, PER-371 and PER-377, were unmethylated according to the msPCR, as evidenced by a positive amplification with the USP primer pair only, shown in Figure 6.3C. In contrast, the cell lines that did not express *CTGF*, PER-485, PER-490, PER-495 and JURKAT, were amplified by the MSP primer pair indicating they were methylated at the MSP priming sites.

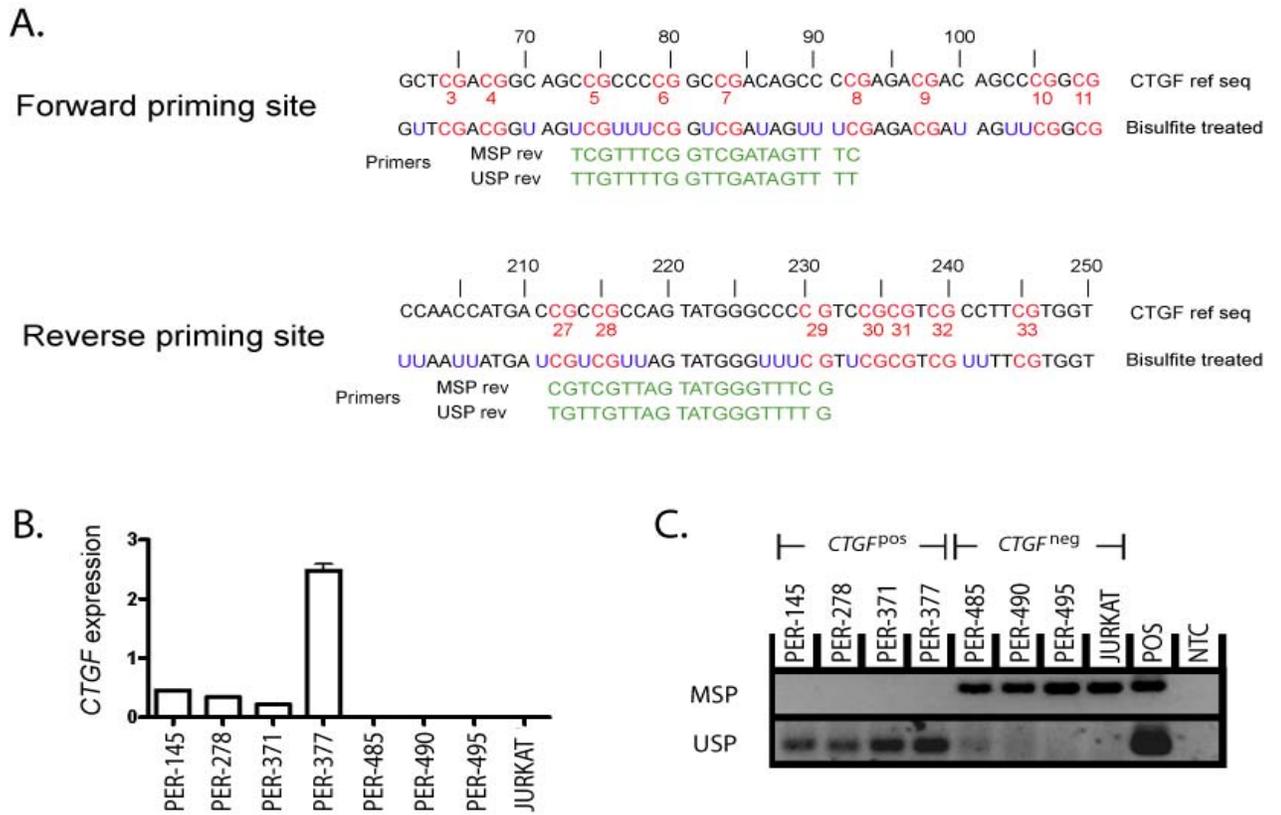


Figure 6.3 Methylation-specific PCR analysis of ALL cell lines

A. *CTGF* reference sequence (ref seq) is shown with nucleotide position relative to the canonical TSS. Unmethylated cytosines are converted to uracil by bisulfite, shown as blue residues in bisulfite-treated sequence. The msPCR priming sites cover CpGs 5 to 8 (numbered in red) at the forward priming site and CpGs 27 to 29 at the reverse priming site. Specific priming by MSP or USP PCR primers (green) infers the methylation status of DNA prior to bisulfite conversion. **B.** *CTGF* expression in a panel of seven B-lineage ALL and one T-ALL cell line, measured by qRT-PCR and normalized to *ACTB*. Error bars indicate the standard error of the mean of technical replicates (n=3). **C.** Cell lines examined by msPCR grouped by *CTGF* mRNA status (positive or negative). For each of the two assays the respective bisulfite-treated positive control was included, as well as a non-template control (NTC). For the USP primers, the BAC clone RP11-69I8 which contains the *CTGF* locus was used while for the MSP primers, the RP11-69I8 BAC was methylated using the SSI methyltransferase prior to bisulfite-treatment.

6.2.3 Bisulfite sequencing of B-lineage ALL cell lines

While the msPCR experiments suggested an association between CpG methylation and reduced *CTGF* expression, a drawback of msPCR is that it restricts examination of the DNA methylation status to only those CpGs targeted by the msPCR primers. Therefore, to gain an understanding of DNA methylation across the entire *CTGF* CpG island, bisulfite sequencing was performed on five of the B-lineage ALL cell lines, three *CTGF*^{pos} (PER-145, PER-278, PER-377) and two *CTGF*^{neg} (PER-485, PER-495). The CpG island was examined as three overlapping amplicons as reported previously (Kikuchi et al. 2007). Amplicon 1 contained nucleotides -113 to +44, Amplicon 2 contained +25 to +566 and Amplicon 3 contained +547 to +916 (numbering relative to TSS). These PCR amplicons were generated from bisulfite-treated genomic DNA, then gel extracted, cloned and sequenced. For each amplicon, eight clones were sequenced in order to gain a representative snapshot of CpG methylation. The results of bisulfite sequencing are shown in Figure 6.4. Consistent with the results of the msPCR, the *CTGF* CpG island was found to be hypomethylated in those cell lines that aberrantly expressed *CTGF* mRNA. The cell line with the highest *CTGF* expression, PER-377, was virtually devoid of CpG methylation at the *CTGF* locus. There was some residual methylation at the 3' end of the CpG island in the lower *CTGF*-expressing cell lines PER-145 and PER-278, with PER-278 showing more pronounced methylation at CpG number 98. In contrast to these findings, the two *CTGF*-negative cell lines PER-485 and PER-495, displayed a high degree of CpG methylation across the *CTGF* locus. This was more pronounced in PER-495, which has only one copy of *CTGF* (see Chapter 4), where the locus was completely methylated.

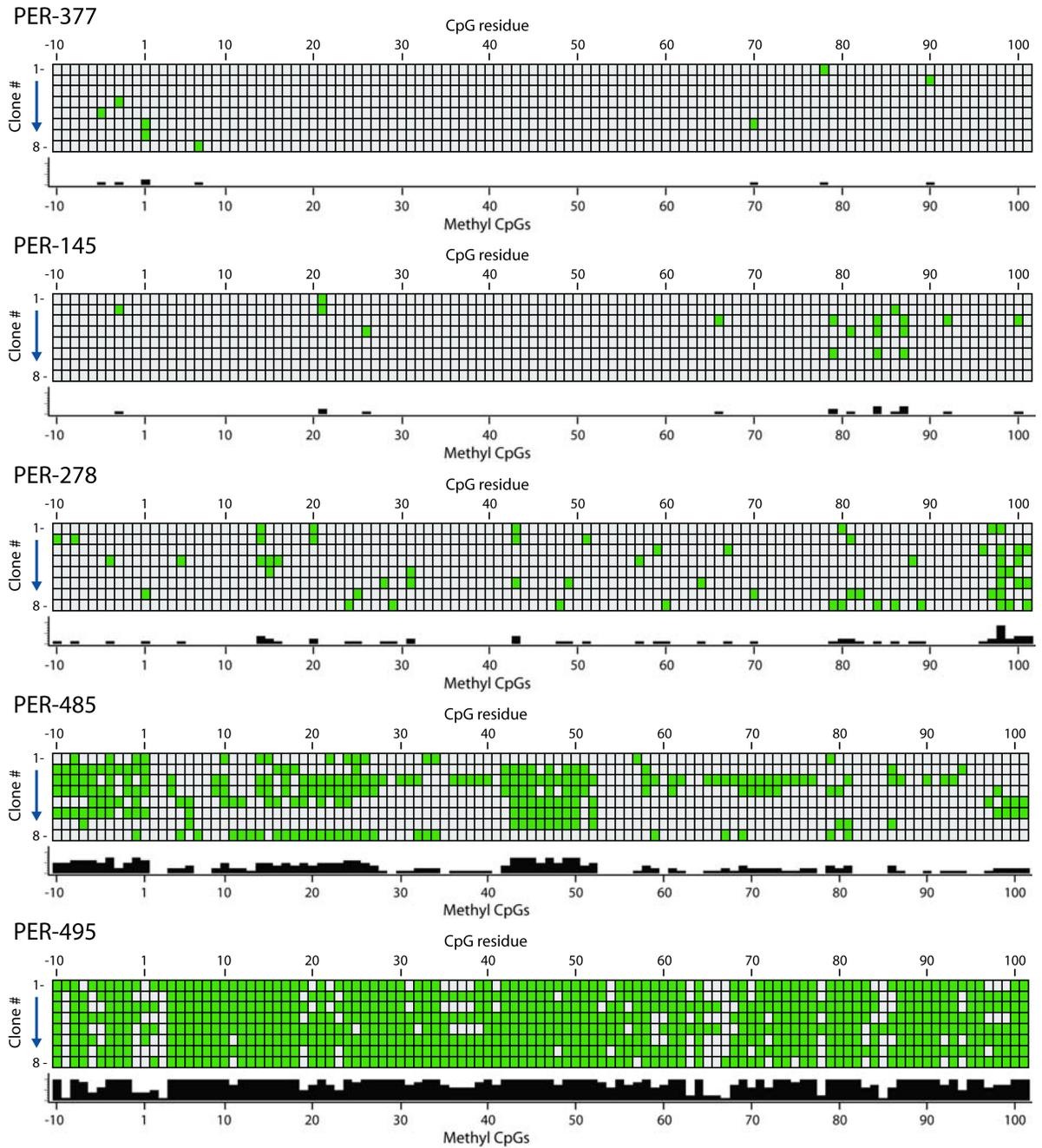


Figure 6.4 Methylation status determined by bisulfite sequencing of the *CTGF* locus correlates with *CTGF* expression in B-lineage ALL cell lines (Figure legend over page)

Figure 6.4 Methylation status determined by bisulfite sequencing of the *CTGF* locus correlates with *CTGF* expression in B-lineage ALL cell lines

The five cell lines examined by bisulfite sequencing were PER-377 (*CTGF*^{high}), PER-145 (*CTGF*^{low}), PER-278 (*CTGF*^{low}), PER-485 (*CTGF*^{neg}) and PER-495 (*CTGF*^{neg}). For each diagram, cell line name is indicated in the top left. Clone numbers 1 to 8 run from top to bottom for each amplicon. The three bisulfite sequencing PCR amplicons are shown as one block from left to right encompassing CpGs -10 to +1, +2 to +78 and +79 to +101 respectively. Methylated CpGs are represented by shaded boxes. The incidence of methylation at each CpG residue is represented in the histogram at the base of each diagram.

6.2.4 Hypomethylation of the *CTGF* locus in primary pre-B ALL

Two independent techniques (msPCR and bisulfite sequencing) demonstrated a correlation between the degree of CpG methylation and expression of *CTGF* in a panel of B-lineage ALL cell lines. To determine if this relationship also existed in primary patient specimens, a panel of pre-B leukaemic bone marrow samples was selected for methylation analysis. Eight samples were analysed using the same msPCR approach that was used to screen the B-lineage ALL cell lines. Of these eight primary samples, qRT-PCR analysis of *CTGF* expression revealed that five exhibited low to negative *CTGF* mRNA expression (*CTGF*^{low/neg}), while three were positive for *CTGF* mRNA at varying levels, collectively referred to herein as *CTGF*^{high} (Figure 6.5A). Despite the heterogeneous levels of *CTGF* expression, all of the primary specimens were amplified by the USP assay only, and thus were classed as unmethylated (Figure 6.5B).

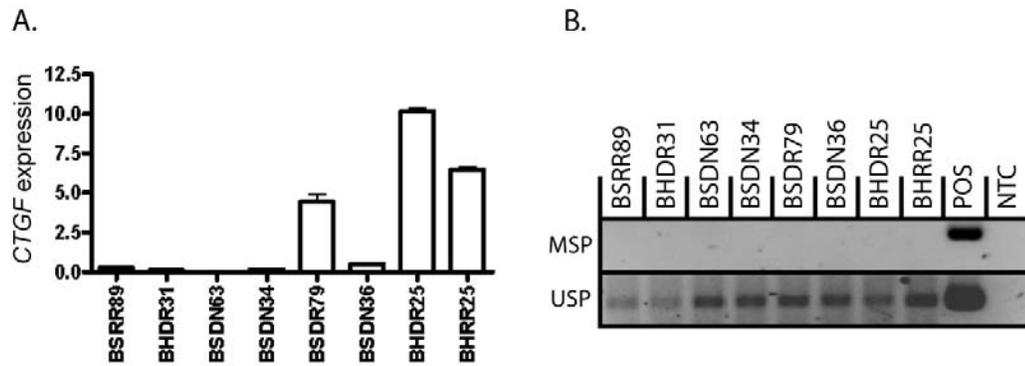


Figure 6.5 msPCR analysis of primary pre-B ALL specimens

A. *CTGF* expression in a panel of primary paediatric pre-B-ALL specimens measured by qRT-PCR and normalized to *ACTB*. Error bars indicate the standard error of the mean of technical replicates (n=3). **B.** Analysis of the methylation status of primary specimens by msPCR. For each of the two assays the respective bisulfite-treated positive control was included, as well as a non-template control (NTC). For the USP primers, the BAC clone RP11-69I8 which contains the *CTGF* locus was used while for the MSP primers, the RP11-69I8 BAC was methylated using the SSI methyltransferase prior to bisulfite-treatment.

6.2.5 Bisulfite sequencing of primary pre-B ALL specimens

The msPCR assay indicated that all eight primary B-ALL specimens tested were unmethylated at the msPCR priming sites, regardless of *CTGF* expression status. However, the methylation status of the rest of the CpG island was unclear. Bisulfite sequencing was therefore undertaken using a larger panel of 14 primary specimens exhibiting varying levels of *CTGF* expression according to microarray data obtained previously (Boag et al. 2007). *CTGF* mRNA levels were validated using qRT-PCR (shown in Figure 6.6). Six of the samples displayed high levels of *CTGF* mRNA (*CTGF*^{high}). The remaining eight samples were classed as *CTGF*^{low/neg}, as seven samples displayed extremely low levels of *CTGF* mRNA and one sample, BSDN63 was completely negative for *CTGF* mRNA after 40 cycles of amplification. Bisulfite sequencing data for the *CTGF*^{high} specimens is shown in below in Figure 6.7 while the *CTGF*^{low/neg} samples are shown in Figure 6.8.

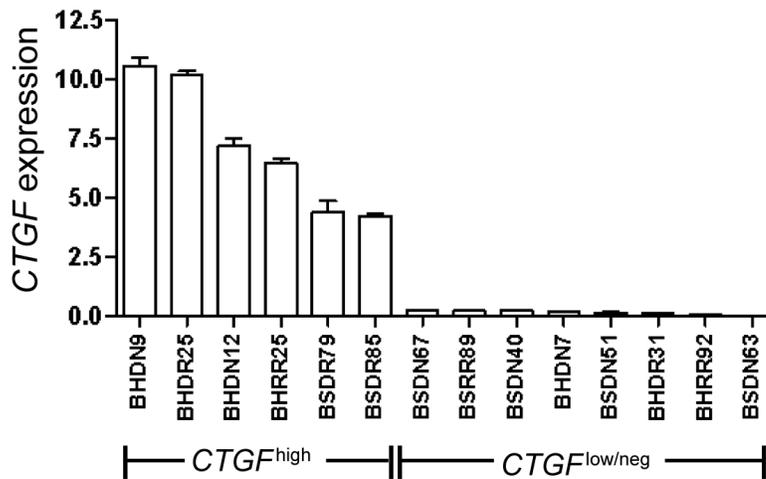


Figure 6.6 *CTGF* mRNA expression in primary pre-B ALL specimens used for bisulfite sequencing

CTGF expression was measured in primary pre-B ALL specimens by qRT-PCR normalized to *ACTB*. Error bars represent the standard error of the mean of technical replicates (n=3). Samples were classed as either $CTGF^{high}$ or $CTGF^{low/neg}$.

Bisulfite sequencing revealed that all of the $CTGF^{high}$ specimens were hypomethylated across the *CTGF* CpG island. However, in contrast to the inverse correlation observed in cell lines between CpG methylation and *CTGF* expression, this correlation was not found in primary specimens, as none of the $CTGF^{low/neg}$ primary specimens displayed significant levels of DNA methylation across the *CTGF* locus. One exception was BHDN7 which displayed methylation of CpGs 15 to 19, 21 to 26, 31 to 33 and 36 to 44 in two clones, suggesting mono-allelic methylation of these CpG residues. There was some sporadic methylated CpGs in BHDR31 which was methylated at the 3' end of Amplicon 3 and BHRR92 was methylated at CpG 24, however no single CpG residue was consistently methylated amongst replicates, indicating that CpG methylation was unlikely to contribute to silencing of *CTGF* expression in these specimens. These results indicate that hypomethylation of the *CTGF* locus is a common feature of primary pre-B ALL specimens, irrespective of *CTGF* expression status.

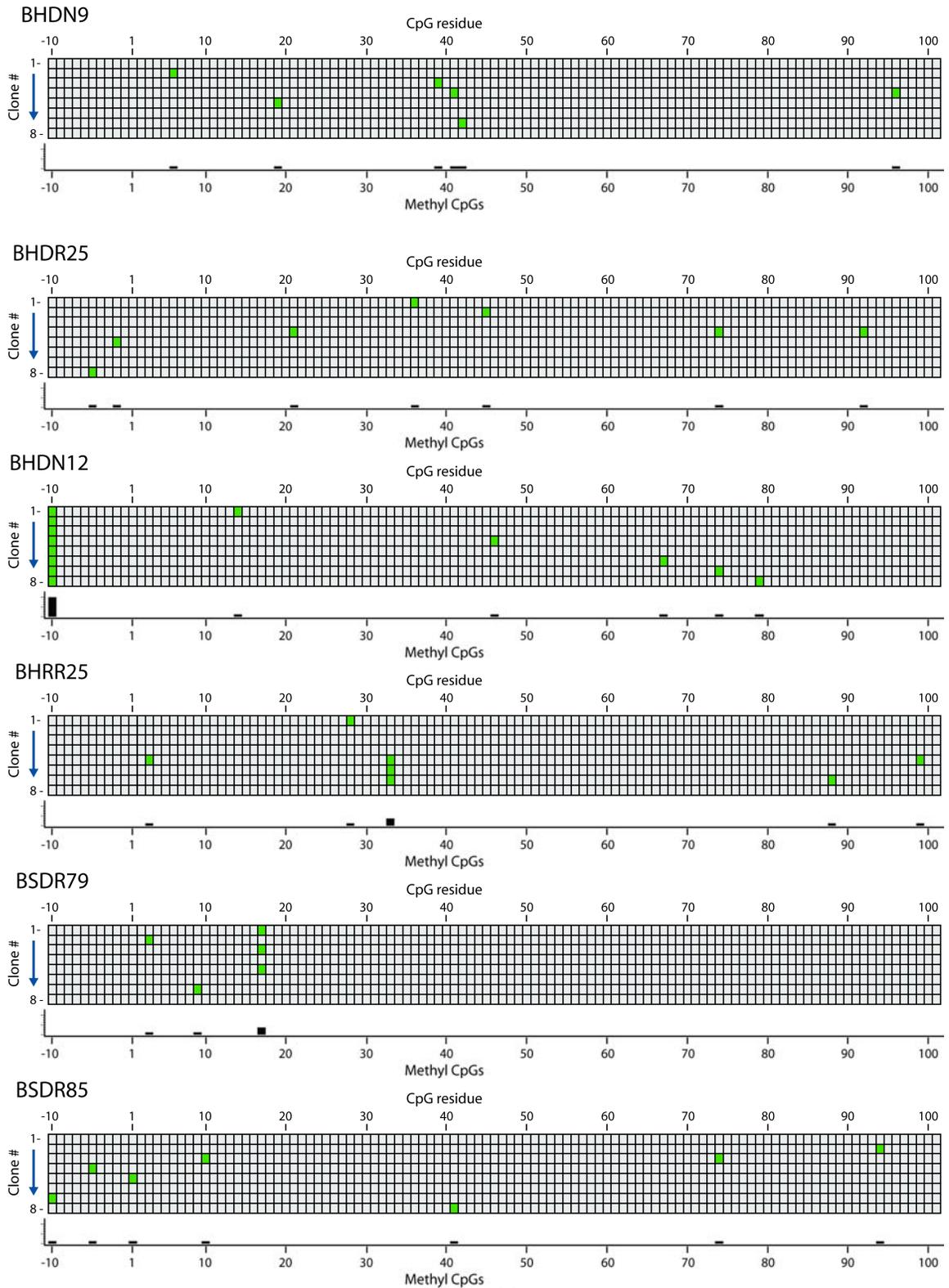


Figure 6.7 Bisulfite sequencing of *CTGF*^{high} primary pre-B ALL specimens (Figure legend over page).

Figure 6.7 Bisulfite sequencing of *CTGF*^{high} primary pre-B ALL specimens
(Continued from previous page)

For each diagram, specimen name is indicated in the top left. Clone numbers 1 to 8 run from top to bottom for each amplicon. The three bisulfite sequencing PCR amplicons are shown as one block from left to right encompassing CpGs -10 to +1, +2 to +78 and +79 to +101 respectively. Methylated CpGs are represented by shaded boxes. The incidence of methylation at each CpG residue is represented in the histogram at the base of each diagram

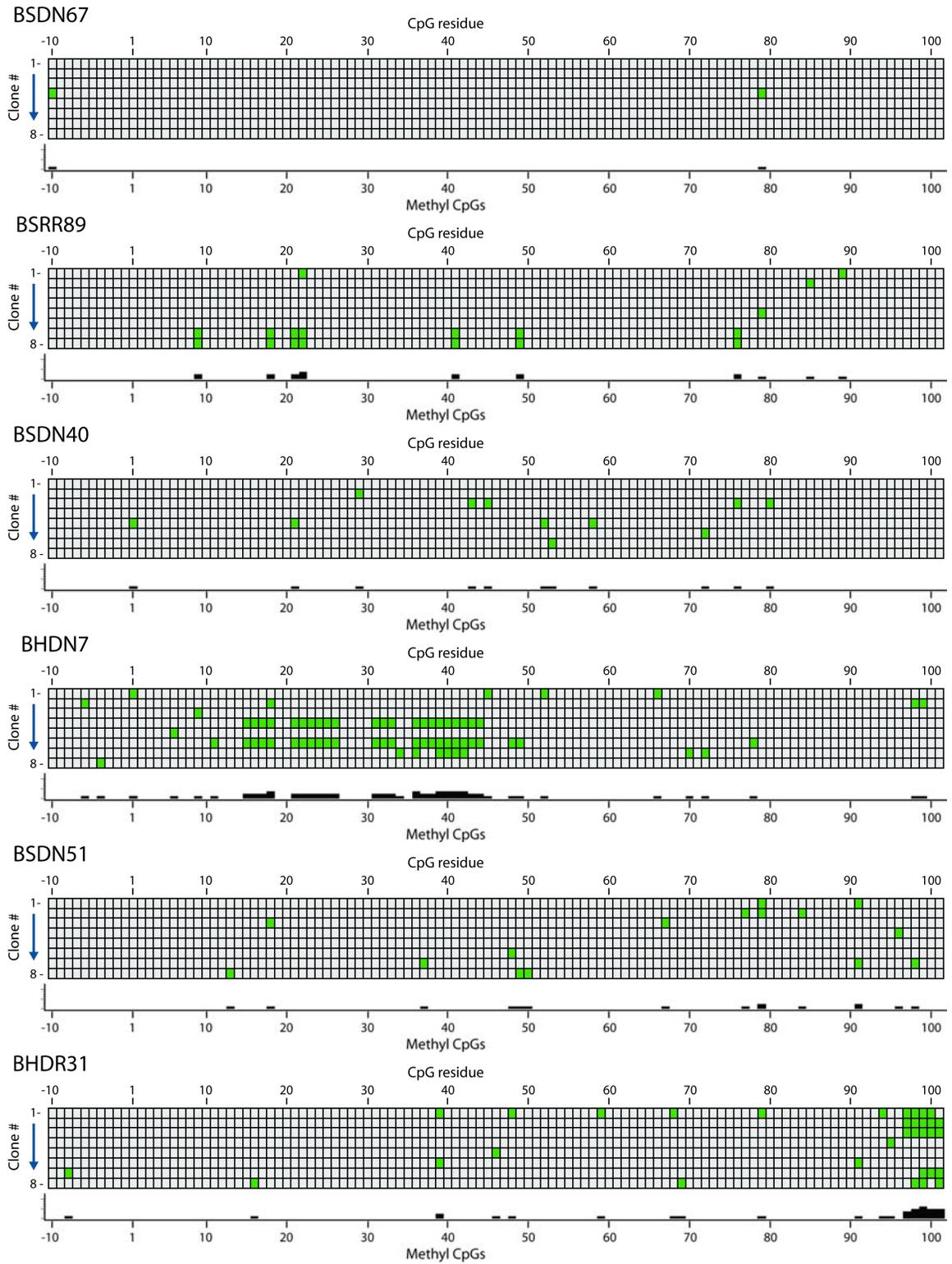


Figure 6.8 Bisulfite sequencing of *CTGF*^{low/neg} primary pre-B ALL specimens

(Figure and legend continued over page).

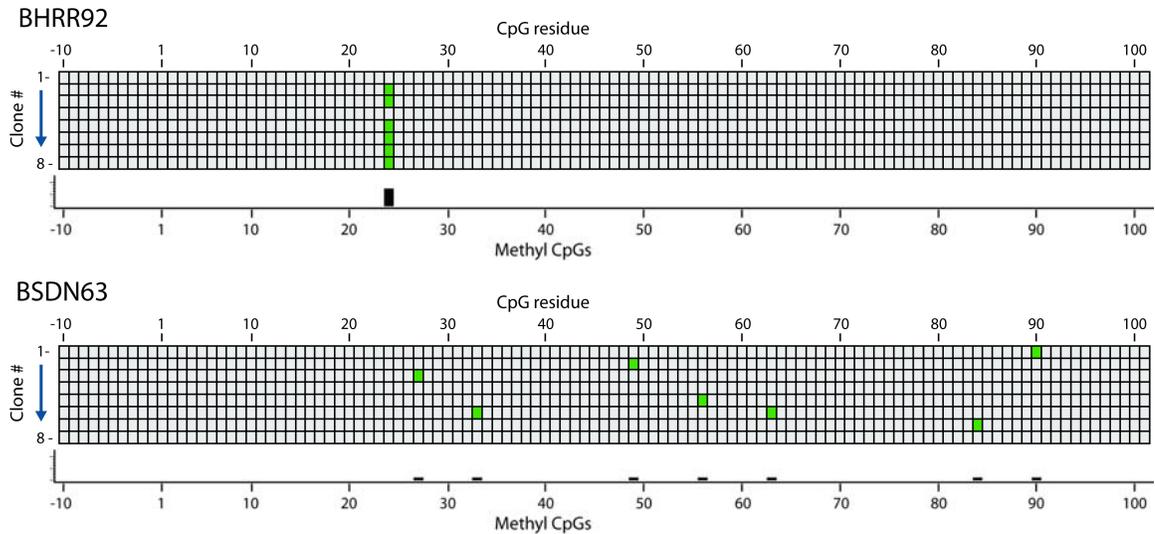


Figure 6.8 Bisulfite sequencing of *CTGF*^{low/neg} primary pre-B ALL specimens
(Continued from previous page)

For each diagram, specimen name is indicated in the top left. Clone numbers 1 to 8 run from top to bottom for each amplicon. The three bisulfite sequencing PCR amplicons are shown as one block from left to right encompassing CpGs -10 to +1, +2 to +78 and +79 to +101 respectively. Methylated CpGs are represented by shaded boxes. The incidence of methylation at each CpG residue is represented in the histogram at the base of each diagram

6.2.6 Bisulfite sequencing of primary T-ALL and CD34^{pos} specimens

Microarray experiments had previously revealed that *CTGF* was not expressed in T-ALL specimens, nor CD34^{pos} cells enriched from normal bone marrow (Boag et al. 2007; Boag et al. 2006). Analysis of T-ALL specimens by qRT-PCR did not detect *CTGF* mRNA (data not shown), and RNA was not available from CD34^{pos} cells. To examine the methylation status of the *CTGF* locus in these two cell types, DNA from two T-ALL primary bone marrow specimens (THDN2 and THDN6) and the CD34^{pos} cells isolated by Boag and colleagues (Boag et al. 2007), were analysed by bisulfite sequencing of the *CTGF* locus. Figure 6.9 shows the results of bisulfite sequencing of these primary specimens. Both of the T-ALL specimens displayed extensive methylation in Amplicons 1 and 3, while the middle amplicon was not consistently methylated, but

did show focal mono-allelic methylation in both cases. Bisulfite sequencing of THDN2, showed that Amplicon 2 was sporadically methylated at different regions on each allele. In THDN6, Amplicon 2 was hypermethylated on one allele, while the other exhibited methylation of a single CpG dinucleotide (CpG 33). Similar to the primary pre-B ALL specimens, CD34-positive bone marrow cells were hypomethylated across all three bisulfite sequencing amplicons.

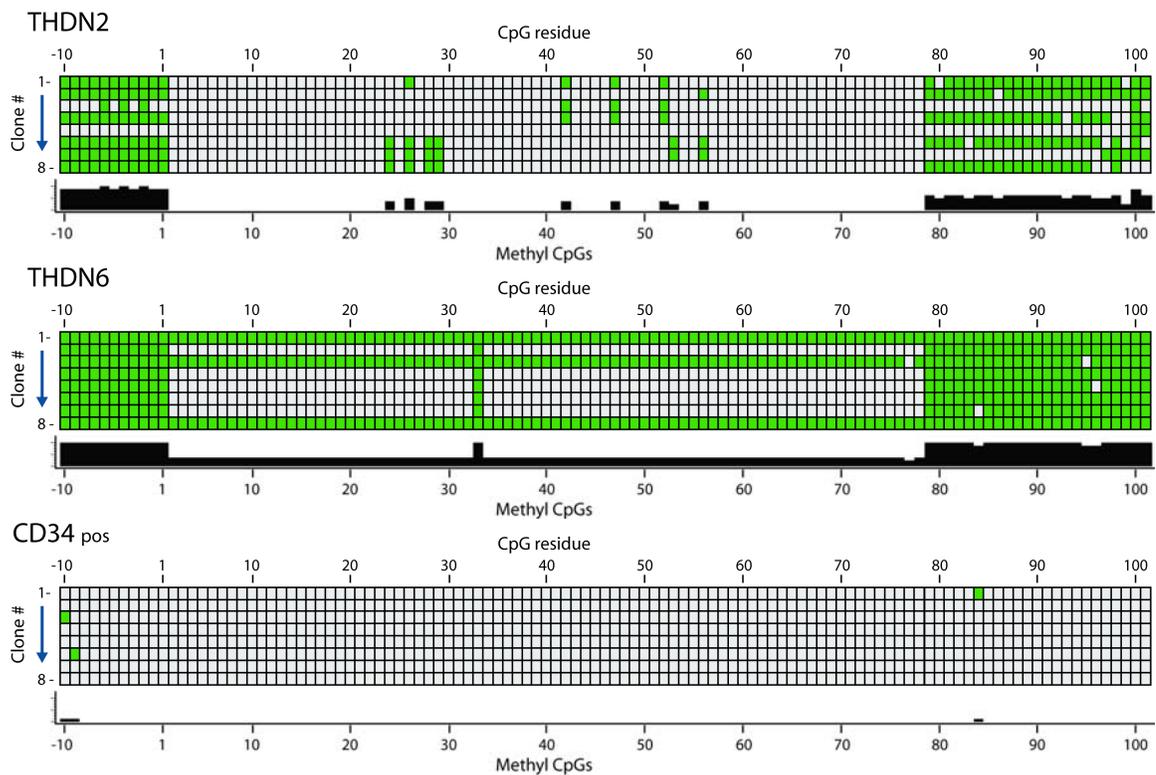


Figure 6.9 Bisulfite sequencing of primary T-ALL specimens and normal CD34 positive bone marrow cells

For each diagram, specimen name is indicated in the top left. Clone numbers 1 to 8 run from top to bottom for each amplicon. The three bisulfite sequencing PCR amplicons are shown as one block from left to right encompassing CpGs -10 to +1, +2 to +78 and +79 to +101 respectively. Methylated CpGs are represented by shaded boxes. The incidence of methylation at each CpG residue is represented in the histogram at the base of each diagram.

6.2.7 Modulation of the global epigenetic state of B-lineage ALL cell lines

Expression of *CTGF* in B-lineage ALL cell lines was highly correlated with the epigenetic status of the *CTGF* locus. Thus, altering the global epigenetic state of ALL cell lines *in vitro* might be expected to lead to changes in *CTGF* gene expression. To test this hypothesis, five B-lineage ALL cell lines were treated with the DNA demethylating agent 5-aza-2'-deoxycytidine (Aza) and the histone deacetylase (HDAC) inhibitor Trichostatin-A (TSA). The three *CTGF*^{neg} cell lines used for this experiment were PER-485, PER-490 and PER-495. In all three cases, treatment with Aza or TSA was highly cytotoxic compared with medium controls and this persisted for two biological replicates (data not shown). Two of these cell lines, PER-490 and PER-485 were derived from infant ALL specimens exhibiting t(4;11) MLL rearrangements which are known to exhibit a marked sensitivity to HDAC inhibitors and Aza (Tonelli et al. 2006; Niitsu et al. 2001). For this reason, these cell lines were not examined further. Therefore two *CTGF*^{pos} cell lines were examined for their response to Aza and TSA, namely PER-145 and PER-377. These two cell lines were shown to be hypomethylated at the *CTGF* locus by bisulfite sequencing, see Figure 6.4.

In the B-ALL cell line PER-377, Aza alone resulted in a modest but not statistically significant increase in *CTGF* expression (Figure 6.10). Treatment with TSA alone also had little effect on *CTGF* expression. However, co-treatment with both agents resulted in a marked and significant increase in *CTGF* expression compared to untreated cells. In the pre-B ALL cell line PER-145, treatment with either agent alone and in combination resulted in a significant increase in *CTGF* expression compared to medium only controls confirming that both DNA methylation and histone acetylation can influence *CTGF* expression in this cell line.

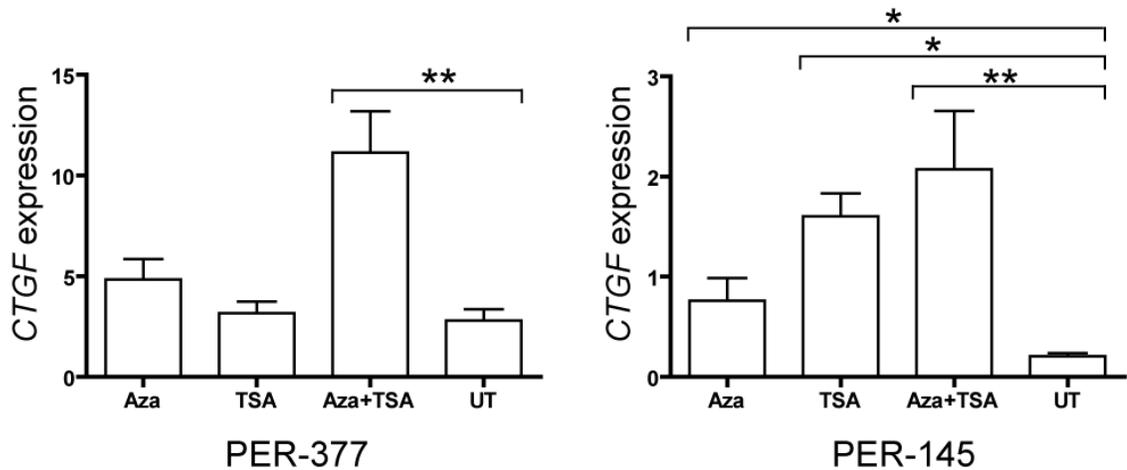


Figure 6.10 The effect of Aza and TSA on *CTGF* expression in B-lineage ALL cell lines PER-377 and PER-145

Cells were treated with 5 μ M 5'aza-2'-deoxycytidine (Aza), 330nM Trichostatin-A (TSA), a combination of both Aza and TSA or culture medium only (UT). Changes in *CTGF* gene expression were measured by qRT-PCR normalized to *ACTB*. Error bars represent the standard error of the mean of three independent experiments. * $p < 0.05$, ** $p < 0.005$.

6.3 Discussion

Investigation of the human genomic *CTGF* locus revealed the presence of a CpG island of approximately 1 kb that was predominantly localised to the 5' end of the *CTGF* gene and coding region. This study showed for the first time that methylation across this region was inversely correlated with *CTGF* gene expression in B-lineage ALL cell lines, thus indicating that this locus can be regulated epigenetically and that hypomethylation, rather than mutation, provides a plausible mechanism for its deregulated expression in pre-B ALL. This finding extended other studies that made use of cell lines derived from hepatocellular carcinoma and ovarian cancers (Chiba et al. 2005; Kikuchi et al. 2007). Several studies have demonstrated that the methylation patterns observed in cancer cell lines closely resemble those observed in the primary tumours from which they were derived (Ueki et al. 2002; Markl et al. 2001; Lind et al. 2004). Furthermore, studies from this laboratory have shown that cell line models can serve as excellent surrogates for

primary tumours (Beesley et al. 2006; Kees et al. 2003). However, novel findings should, where possible, be validated in primary tumour specimens.

6.3.1 The *CTGF* locus is hypomethylated in primary pre-B ALL

Bisulfite sequencing of primary patient specimens provided a detailed map of CpG methylation of the *CTGF* locus in paediatric pre-B ALL specimens and confirmed that hypomethylation of the *CTGF* coding region is a general feature of this disease, regardless of the level of *CTGF* expression, which directly contrasted with observations made in the B-lineage ALL cell lines. When selecting primary bone marrow specimens for bisulfite sequencing analysis, specimens exhibiting a range of *CTGF* expression levels were selected using microarray data obtained in previous studies (Boag et al. 2006; Boag et al. 2007). When these primary specimens were re-examined for *CTGF* expression using qRT-PCR, 6 out of 14 pre-B ALL specimens displayed high levels of *CTGF* mRNA, while seven of the remaining eight *CTGF*^{low/neg} specimens exhibited extremely low *CTGF* mRNA levels that were unlikely to result in synthesis of biologically relevant levels of CTGF protein. In contrast, when the two T-ALL primary specimens and *CTGF*^{neg} B-lineage ALL cell lines, PER-485, PER-490 and PER-495 were examined by qRT-PCR, *CTGF* mRNA was completely undetectable after 40 cycles of amplification, and this corresponded to hypermethylation at the *CTGF* locus. RNA was not available from the CD34^{pos} specimen for this study, however they were considered negative for *CTGF* mRNA based on microarray data (Boag et al. 2007). These findings suggest a model where hypomethylation of the *CTGF* locus leaves this region susceptible to promiscuous transcriptional activation in primary pre-B ALL, while hypermethylation of the *CTGF* locus in T-ALL and *CTGF*^{neg} cell lines renders this region transcriptionally silent.

This observation is supported by previous findings that identified decreased methylation at the *CTGF* locus in pre-B ALL specimens compared to T-ALL and DLBCL (Martin-Subero et al. 2009). The assertion that the *CTGF* locus is, by contrast, hypermethylated in T-ALL is also supported by Martin-Subero and colleagues, although analysis of a

greater number of patient specimens is required to confirm this definitively. With the exception of CD34^{pos} cells, normal haematopoietic precursor cells were not available for analysis during this study. However, given the observation that both DLBCL and T-ALL cells were hypermethylated at the *CTGF* locus, and conversely this region was hypomethylation in bone marrow derived CD34^{pos} cells, and pre-B ALL cells, it is conceivable that that epigenetic silencing of the *CTGF* locus may occur during normal haemopoiesis. Careful examination of the methylation status of the *CTGF* locus throughout haemopoiesis should be undertaken to determine if and when epigenetic silencing of the *CTGF* locus occurs.

6.3.2 The importance of histone modifications

The heterogeneity in *CTGF* expression between the *CTGF*^{high} and *CTGF*^{low/neg} primary pre-B ALL specimens may also be a result of histone modifications such as methylation or acetylation which can alter chromatin accessibility and gene expression (Smith & Shilatifard 2011). This hypothesis was supported by the observation that *CTGF* expression was enhanced in response to treatment with the HDAC inhibitor TSA in the PER-145 cell line, which already exhibited moderate *CTGF* expression. Reinforcing this finding, HDAC inhibitors have been shown previously to up-regulate *CTGF* expression in renal epithelial cells, hepatoma cell lines and primary hepatocytes (Komorowsky et al. 2009; Chiba et al. 2004). Care needs to be taken when interpreting these results however, as pharmacological inhibition of HDACs is a blunt instrument in investigating the role of histone acetylation in gene regulation and changes in acetylation cannot be targeted to specific loci. A more detailed analysis of the role of histone acetylation in the regulation of the *CTGF* locus could be undertaken by performing chromatin immunoprecipitation (ChIP) using antibodies against acetylated H3 and H4 histones, and comparing *CTGF*^{high} and *CTGF*^{low/neg} pre-B ALL specimens by ChIP-sequencing or ChIP-on-Chip (Ho et al. 2011). The enormous power of next generation sequencing means that global patterns of histone acetylation in pre-B ALL could be elucidated, and this may prove useful for not just analysis of the *CTGF* locus, but other regions of interest in pre-B ALL.

A commonly held view of the role of CpG islands is one where methylated CpGs can direct repression of gene transcription by recruiting methyl-CpG binding proteins such as the methyl-CpG binding domain (MBD) proteins 1 to 4 to influence gene expression (Feng & Zhang 2001). These proteins prevent gene transcription by inhibiting the binding of transcription factors to DNA regulatory elements or by recruiting histone modifying complexes to alter the local chromatin structure. For example, MBD1 recruits the histone methyltransferase SETDB1 which catalyses tri-methylation of lysine 9 on histone H3 (H3K9me3), which is a repressive mark (Snowden et al. 2002; Sarraf & Stancheva 2004). Recent findings however suggest that CpG islands can also be interpreted through the specific recognition of non-methylated CpGs and furthermore that non-methylated CpG islands can reinforce active chromatin marks within a coding region.

Two important histone modifying enzymes, KDM2A and CFP1, have been shown to specifically bind to non-methylated CpGs through their zinc finger (ZF)-CxxC domain (Blackledge et al. 2010; Lee & Skalnik 2005). KDM2A is a histone demethylase and catalyses the removal of repressive dimethyl marks from H3 lysine 36 (H3K36me2) (Li et al. 2009a; Tsukada et al. 2006). ChIP assays demonstrated that while intergenic regions show constant high levels of H3K36Me2, non-methylated CpG islands occupied by KDM2A were depleted of this repressive mark (Blackledge et al. 2010). CFP1 also contains a ZF-CxxC domain and specifically binds non-methylated CpGs. CFP1 associates with the SET1 complex to catalyse the addition of the trimethyl marks to H3 lysine 4 (H3K4me3) (Lee & Skalnik 2005). H3K4me3 is generally associated with the 5' end of gene coding regions as is strongly associated with transcriptional activation (Okitsu & Hsieh 2007). These data clearly demonstrate that these ZF-CxxC proteins bind to non-methylated DNA and enforce active chromatin marks and encouraging transcription of target genes. Determining the array of histone modifications at the *CTGF* locus may shed light on mechanisms controlling gene expression through alterations in chromatin structure. This is no small task due to the large number of histone modifications that can be examined, however this should be considered a

worthwhile goal for future investigation into the role of epigenetic modification in regulating the *CTGF* locus.

6.3.3 Conclusions

In summary, the experiments described in this chapter demonstrated that the *CTGF* locus contains a CpG island extending from the proximal promoter into exon 3, that is hypomethylated in primary paediatric pre-B ALL and thereby provides a mechanism to explain the tendency of this locus to be deregulated in pre B-ALL. Furthermore, analysis of T-ALL specimens and bone marrow derived CD34^{pos} cells points to the potential developmental methylation (and silencing) of the *CTGF* locus during lymphopoiesis, although more evidence is required before any definitive conclusions can be drawn. Modulation of global methylation and histone acetylation influenced *CTGF* expression in B-lineage ALL cell lines and further investigation into the covalent modification of histones occupying *CTGF* associated nucleosomes is warranted. Such an investigation may reveal differences in chromatin state between pre-B ALL cells exhibiting non-methylated CpG island but different levels of *CTGF* expression.

Chapter 7

General Discussion

Chapter 7

General Discussion

7.1 Introduction

The prognosis for patients diagnosed with ALL is excellent compared to other cancers, achieving cure rates of 85-90% for some forms of this disease, yet ALL continues to be the most frequently diagnosed cancer in children worldwide (Pui et al. 2010; Gaynon et al. 2010; Conter et al. 2010). Moreover, while these cure rates are remarkable, around 20% of children diagnosed with ALL still relapse (Conter et al.; Gaynon et al. 2010; Pui et al. 2010), and improving therapy for these patients is of paramount importance. Advances in long-term survival for ALL patients have come about largely through the careful optimisation of treatment protocols, facilitated by large, multi-centre clinical trials (Gaynon et al. 2010; Pui et al. 2010), and not through the development of novel therapeutic agents. One of the most critical aspects of conventional ALL therapy is stratifying patients based on clinical features present at diagnosis, and directing these patients to the appropriate level of treatment (Pieters & Carroll 2010). The aim is to identify patients at a high risk of relapse, and ensure these patients are directed to intensive chemotherapy, to afford them the best chance of achieving a clinical remission. At the same time, low or standard-risk patients are spared from the more toxic therapies reserved for high-risk patients or those that relapse. These treatments can carry significant long-term health risks, and thus limiting their use in children is desirable (Dowling et al. 2010).

However, a key question remains in terms of what biological features can best distinguish these high-risk patients from others and how therapy can be tailored further for high-risk patients according to specific biological features, such as gene expression signatures or clinical features. Clinical trials alone cannot answer these questions, thus the key to improving patient outcomes further is to elucidate the molecular mechanisms underlying ALL. This may facilitate the discovery of novel prognostic biomarkers with

therapeutic relevance, or the elucidation of tumour specific cellular pathways that can be exploited by new or existing cancer treatments for therapeutic gain.

7.2 Study objectives

This thesis investigated the molecular mechanisms responsible for deregulated expression of the proto-oncogene *CTGF* in childhood pre-B ALL. The literature review highlighted that aberrant *CTGF* expression is a frequent event in pre-B ALL (Vorwerk et al. 2000; Sala-Torra et al. 2007; Boag et al. 2007), as well as numerous solid cancers, however this review also highlighted an incomplete understanding of both the mechanisms regulating the *CTGF* gene locus and its involvement in the pathogenesis of ALL. Hence, this thesis addressed four related research questions. The first hypothesis was that *CTGF* expression may be a result of abnormal transcription factor activation or expression. Therefore, the global gene expression profiles of three, independent, paediatric pre-B ALL cohorts, were compared to identify *CTGF*-correlated patterns of gene expression. The second hypothesis was that *CTGF* expression may exhibit prognostic significance and may be associated with clinical features at the time of the patient's diagnosis. Thus, the association between clinical features, *CTGF* expression and patient survival was investigated in the PMH cohort of paediatric pre-B ALL. The third hypothesis was that aberrant *CTGF* expression may have a genetic basis. To address this, a sequencing and hybridisation approach was employed to investigate the *CTGF* gene locus and *CTGF* mRNA transcripts for lesions that may contribute to its deregulated expression. Lastly, the fourth hypothesis was that *CTGF* expression has an epigenetic basis. Hence, methylation-specific PCRs and bisulfite sequencing was performed to investigate the association between DNA methylation and *CTGF* expression. These experiments were expected to shed light on the mechanisms driving *CTGF* expression in pre-B ALL, as well as highlighting its prognostic relevance.

7.3 Critical findings

7.3.1 A possible role for Ikaros proteins in *CTGF* transcription

To examine whether aberrant transcription factor activation or expression was responsible for transcriptional activation of *CTGF* in pre-B ALL, three independent, pre-B ALL microarray datasets were compared and examined for *CTGF*-correlated genes. Six genes were identified by this approach and by comparing the 5' flanking regions of these six genes to that of *CTGF*, as well as each other, it was expected that common transcription factor binding sites would become evident, implicating individual transcription factors as candidates for functional analysis. Somewhat surprisingly, binding sites for nuclear factors known to regulate the *CTGF* locus in other cell types were not identified in these *CTGF*-correlated genes. Instead, *in silico* analysis uncovered binding sites for the Ikaros protein family members IKAROS and HELIOS, in the 5' flanking sequence of these genes.

The Ikaros family of transcriptional regulators have been associated with both transcriptional activation, as well as silencing of target loci by targeting repressive chromatin changes (Koipally et al. 1999; Rebollo & Schmitt 2003). Both HELIOS and IKAROS have been associated with haematological malignancies and *IKAROS* is frequently deleted or mutated in high-risk pre-B ALL (Kuiper et al. 2010; Mullighan et al. 2009). Thus, the Ikaros family of proteins may be involved in silencing the *CTGF* locus, and a loss of IKAROS or HELIOS function may result in dysregulation of *CTGF* expression. Indeed analysis of a cohort of high-risk pre-B ALL patients that were recently examined for *IKAROS* mutations and deletions (Mullighan et al. 2009), confirmed that *CTGF* expression was significantly associated with *IKAROS* genomic lesions. While these findings suggest that IKAROS and/or HELIOS may play a role in regulating the *CTGF* locus, detailed functional characterisation of the effect of these proteins upon the *CTGF* promoter have not yet been undertaken.

This is the first time IKAROS or HELIOS have been implicated in *CTGF* gene expression. The role of the Ikaros family gene mutations and deletions in dysregulation

of the *CTGF* locus and in pre-B ALL more broadly is warranted. *IKAROS* gene mutations or deletions are highly predictive of relapse (Kuiper et al. 2010; Mullighan et al. 2009), and in-frame deletions within the *IKAROS* coding region can result in the synthesis of inhibitory DN isoforms. However, these may also be generated by alternative splicing (Meleshko et al. 2008), making this a challenging area of research. Compounding this, DN isoforms may interact with other Ikaros family members due to conservation in the dimerisation domain of this protein family. Thus, a broad analysis of the mutation and transcriptional status, as well as the functional capacity of all five Ikaros family members should be performed simultaneously in any future studies and correlated to *CTGF* gene expression.

7.3.2 Clinical importance of *CTGF* expression in pre-B ALL

While *CTGF* expression has been associated with a poor outcome in both adult and high-risk paediatric ALL (Sala-Torra et al. 2007; Kang et al. 2010), it was unclear if specific biological features correlated with *CTGF* mRNA levels. Detailed patient data was available for the PMH cohort, and thus associations between a range of clinically relevant biological features and *CTGF* gene expression was examined. Significantly lower mean *CTGF* expression was observed in patients with enlarged lymph nodes, and there was also a similar trend in patients with enlarged spleens, however this did not reach significance. This observation suggests that *CTGF* expression in leukaemic cells does not promote metastasis to sites outside of the bone marrow. As reported in Chapter 1, *CTGF* expression has been both positively and inversely associated with metastasis (Lin et al. 2005; Li et al. 2007), and thus the tissue context is likely to significantly influence the biological outcome of *CTGF* secretion.

To investigate the associations between *CTGF* expression and patient outcomes, the PMH cohort was stratified based upon *CTGF* expression and examined for RFS and overall survival. There was no association between overall survival and *CTGF* expression, however those patients whose leukaemic blasts were *CTGF*^{POS} had a lower RFS compared to those patients whose blasts were *CTGF*^{low/eg} (71% and 83%

respectively). This association in the PMH cohort did not reach statistical significance, and this may be due to the cohort composition which contained 68% standard-risk and 32% high-risk patients. Nonetheless, these findings confirm reports by others that *CTGF* expression is associated with a poor outcome in pre-B ALL. The mechanisms through which *CTGF* exerts its biological effects in ALL remains unclear, however it is suspected that aberrant *CTGF* gene expression in the bone marrow may result in changes in the microenvironment that create a preferred site of sanctuary for leukaemic cells, and this may have important ramifications in regards to the emergence of drug resistance during therapy.

7.3.3 A model for CTGF-mediated leukaemogenesis

The *CTGF* protein is pleiotropic, however *CTGF* has no known role in lymphocyte biology. It is expressed in numerous tumour types, and thus may confer a biological advantage to leukaemic cells through autocrine or paracrine mechanisms, or by altering interactions with the stromal microenvironment. It is now widely accepted that bone marrow stromal cells facilitate adhesion of leukaemic cells within the bone marrow microenvironment, affording them sanctuary from chemotherapy (Weisberg et al. 2008; Meads et al. 2008). *CTGF* can induce terminal differentiation of bone marrow stromal cells (BMSCs) into fibroblasts (Lee et al. 2010) and these BMSC-derived fibroblasts may exhibit some of the growth-promoting effects exhibited by cancer-associated fibroblasts in other cancer types (Allen & Louise Jones 2011).

This proposed model, where secreted *CTGF* has a paracrine effect upon the bone marrow microenvironment is outlined in Figure 7.1, and is supported by two key unpublished findings generated in the laboratory of Prof Ursula Kees. Firstly, both *CTGF*^{pos} and *CTGF*^{neg} pre-B ALL cells fail to respond to treatment with exogenous recombinant human *CTGF* protein *in vitro*, seemingly ruling out autocrine growth effects upon these leukaemic cells. Secondly, pre-treatment of human bone marrow stromal cells (HS5 cell line) with rh*CTGF* resulted in enhanced adhesion, migration, proliferation and most notably drug-resistance, of pre-B ALL cells in subsequent co-

culture experiments. It could be speculated that a paracrine growth loop exists that is facilitated by CTGF-induced changes in stromal cell gene expression. These unpublished findings, which were not part of the present study, may account for the enhanced bone marrow fibrosis observed in pre-B ALL (Noren-Nystrom et al. 2008). This model also provides a plausible mechanism by which CTGF expression in the bone marrow may promote drug resistance and relapse in patients, thus highlighting the importance of better understanding the mechanisms driving its deregulated expression. These studies are ongoing and are expected to yield important data regarding the paracrine effects of CTGF in the bone marrow microenvironment.

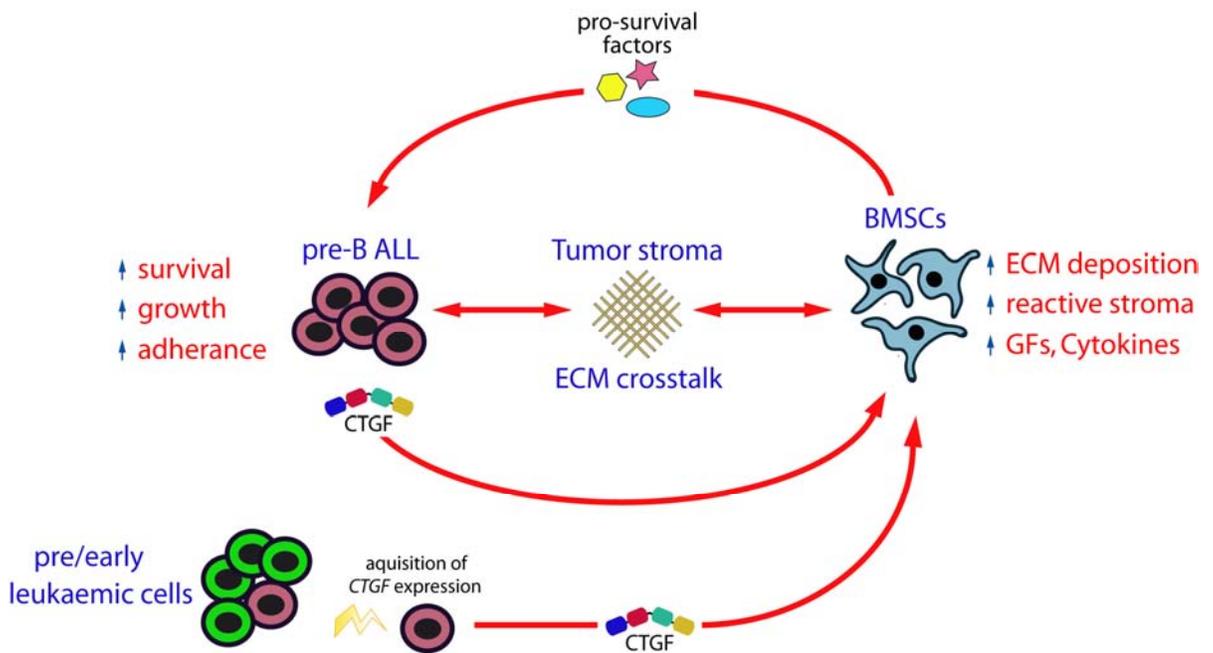


Figure 7.1 Proposed model of the growth-promoting effects of CTGF in the bone marrow microenvironment

Acquisition of *CTGF* gene expression by pre-leukaemic or early leukaemic cells results in aberrant secretion of CTGF protein. Bone marrow stromal cells (BMSCs) respond to CTGF protein by releasing pro-survival signals such as growth factors and cytokines as well as increasing synthesis of extracellular matrix (ECM) proteins. BMSC derived pro-survival factors promote enhanced proliferation and adherence of pre-B ALL cells. Tumour stroma may also facilitate crosstalk between ALL cells and other stromal cells in the bone marrow microenvironment.

7.3.4 Genomic lesions did not contribute to *CTGF* expression

The third major hypothesis of this study was that aberrant *CTGF* expression was the result of genomic lesions affecting the *CTGF* locus. Southern blotting ruled out rearrangement of a 16 kb region encompassing the *CTGF* locus and flanking regions. Similarly, gene copy number analysis revealed that primary pre-B ALL specimens and ALL cell lines were diploid for *CTGF*. Thus gross structural alterations could not account for abnormal *CTGF* expression. Sequencing of the *CTGF* promoter and 3'UTR in ALL cell lines did not identify any mutations that could account for either abnormal promoter activation, or a failure in post-transcriptional regulation of *CTGF* mRNA. It is possible that cryptic enhancer or promoter elements outside of the region sequenced in this study may be responsible for the activation of the *CTGF* locus, and sequencing of upstream regions may yet uncover mutations relevant to *CTGF* expression. It should be noted that the rs6918698 C>G SNP at position -739 was identified in one of the *CTGF*^{pos} ALL cell lines, however given the fact that it was only identified in one of four *CTGF*^{pos} cell lines, it seems unlikely that this SNP contributes significantly to *CTGF* promoter dysregulation. Nonetheless, investigation of the incidence of this SNP in primary pre-B ALL specimens may reveal association with *CTGF* expression, and thus could be considered in future studies if only to rule out its involvement.

7.3.5 *CTGF* mRNA is subject to alternative splicing in pre-B ALL

While evidence of aberrant *CTGF* expression has been provided from several sources (Sala-Torra et al. 2007; Vorwerk et al. 2000; Boag et al. 2007), these findings have all been generated by detection of *CTGF* mRNA using qRT-PCR and microarrays, targeting the 3' UTR of the *CTGF* transcript. Northern blotting and cDNA sequencing was used to evaluate the complement of *CTGF* mRNA transcripts produced in ALL cell lines. These experiments confirmed that while the majority of *CTGF* mRNA (>90%) was the canonical 2.4 kb mRNA, an array of *CTGF* transcripts were generated by alternative pre-mRNA splicing, as well as alternative transcription start site usage in ALL cells. This is the first evidence of alternative splicing of *CTGF* mRNA in any tissue type and thus is an important finding, as some of these *CTGF* spliceforms may exist in

other cancer types or normal tissues and have the potential to encode truncated CTGF proteins as well as synthesizing transcripts that can avoid post-transcriptional control of CTGF expression. Analysis of *CTGF* mRNA splicing in primary pre-B ALL specimens was not performed in the present study. However, such an investigation is warranted to establish accurate patterns of *CTGF* mRNA synthesis in pre-B ALL that may uncover tumour-specific *CTGF* transcripts with altered coding potential or susceptibility to post-transcriptional control mechanisms.

7.3.6 *CTGF* mRNA may not reflect CTGF protein levels in ALL

A disparity was observed between *CTGF* mRNA and protein levels in ALL cell lines. While four cell lines were *CTGF*^{pos}, only the PER-377 cell line exhibiting the highest level of *CTGF* mRNA synthesized detectable levels of CTGF protein, suggesting that post-transcriptional regulation of *CTGF* mRNA may be occurring in these cell lines. This finding suggests that the levels of *CTGF* mRNA detected in primary specimens may not accurately represent the amount of CTGF protein produced by these cells. Therefore, CTGF protein would make a more attractive target as a biomarker and may hold more prognostic significance than mRNA levels alone. Future studies should investigate the correlation between *CTGF* mRNA and protein levels in primary specimens. Furthermore, evaluation of the association between CTGF protein expression and patient outcomes is justified.

If CTGF protein levels are to be examined for prognostic significance in pre-B ALL, this raises the question as to what biological material is most appropriate. Diagnostic bone marrow aspirates are routinely collected to confirm a diagnosis of ALL. These specimens represent the most relevant biological material for detecting CTGF protein levels, as CTGF is predicted to significantly alter the bone marrow microenvironment. Purification of mononuclear cells from diagnostic bone marrow aspirates is usually performed with a ficoll-hypaque density gradient. This protocol could be easily adapted to ensure that the acellular plasma fraction of this density gradient is retained and cryopreserved for subsequent molecular investigation of secreted growth factors present

in the bone marrow space. This method would fit into current diagnostic sample processing protocols without significant procedural changes, however aspirates would need to be obtained using the anticoagulant ethylenediaminetetraacetic acid (EDTA) only, as the other commonly used anticoagulant heparin would bind to CTGF through its CT domain, potentially confounding protein detection. Collection of this material for retrospective analysis should be encouraged to focus not only on CTGF, but also other biomarkers present in the bone marrow that may come to light in future research.

7.3.7 The *CTGF* locus is hypomethylated in pre-B ALL

The final hypothesis of this study was that aberrant *CTGF* expression may be a result of epigenetic changes in ALL cells. DNA methylation and histone acetylation has been shown to play a role in modulating the transcriptional potential of the *CTGF* locus in liver and ovarian cancers (Chiba et al. 2004; Chiba et al. 2005; Kikuchi et al. 2007) and global hypomethylation of cancer cells is an accepted mechanism by which oncogenes can be activated in cancer cells (Feinberg & Vogelstein 1983b; Tsujiuchi et al. 1999; Hanada et al. 1993). A combination of msPCR and bisulfite sequencing was used to investigate DNA methylation at the *CTGF* locus in ALL cell lines and primary specimens. This was the first such detailed analysis of epigenetic regulation of the *CTGF* locus in ALL.

An inverse correlation between *CTGF* gene expression and DNA methylation was observed in a panel of ALL cell lines. In contrast, all fourteen pre-B ALL primary specimens were hypomethylated at the *CTGF* locus despite displaying heterogeneous levels of *CTGF* expression. It is not clear whether the *CTGF* gene locus is normally subjected to epigenetic silencing during lymphoid specification or maturation, however in contrast to pre-B ALLs the *CTGF* locus was hypermethylated in two T-ALL specimens investigated in this study and this finding is supported by others that have observed a similar pattern in global methylation analysis of a large panel of haematological malignancies (Martin-Subero et al. 2009). Thus, hypomethylation of the locus is a conceivable mechanism by which aberrant *CTGF* expression may arise. The

ability of epigenetic factors, including DNA methylation and histone acetylation to modulate *CTGF* expression were confirmed *in vitro*, and this finding is also in agreement with published data from others relating to epigenetic regulation of *CTGF* in other cancer types (Kikuchi et al. 2007; Chiba et al. 2004). Because *CTGF* mRNA is not detectable at any stage in normal haemopoiesis, it is highly likely that this locus is epigenetically silenced during normal development. Careful analysis of the epigenetic state of the *CTGF* locus at each stage of lymphopoiesis is a worthwhile endeavour to confirm whether epigenetic processes are responsible for maintaining this region in a transcriptionally silent state.

7.4 Future directions

7.4.1 Prognostic relevance of CTGF

CTGF mRNA was associated with lower RFS in the PMH cohort of paediatric ALL and unpublished data implicates *CTGF* protein in modulating the bone marrow microenvironment to support leukaemic cells. The prognostic relevance of secreted *CTGF* protein in the bone marrow should be investigated, as there is abundant data supporting associations between aberrant *CTGF* mRNA expression and unfavourable outcome in both paediatric and adult ALL. As discussed, diagnostic bone marrow specimens represent the most appropriate biological material for this analysis as this material would provide a snapshot of protein levels in the bone marrow space. Thus, bio-banking of this material should be encouraged to facilitate retrospective analysis of the associations between patient outcomes and secreted growth factors, including *CTGF* or other factors that are highlighted as important candidates in the pathogenesis of ALL.

7.4.2 Modeling the role of CTGF in ALL pathogenesis in-vivo

Murine models of B-lineage lymphomas or leukaemias, as well as xenografts of primary human pre-B ALLs represent suitable *in vivo* models with which to study the effect of *CTGF* gene expression upon the biology of leukaemia, as well as the bone

marrow microenvironment. Human pre-B ALL murine xenographs provide the opportunity to examine whether inhibition of CTGF activity can affect the viability, engraftment or drug resistance profiles of pre-B ALL cells *in vivo*. In an orthotopic murine model of pancreatic adenocarcinoma, in which *CTGF* is expressed at high levels, administration of a human monoclonal CTGF antibody attenuated tumour growth and metastasis. This study demonstrated that secreted CTGF can have a significant effect upon tumour biology and the surrounding stroma, and further that blocking the function of CTGF can influence tumour growth (Aikawa et al. 2006). This highlights the potential for a similar role in ALL therapy. Preliminary studies have shown that co-incubation of pre-B ALL cells with CTGF-treated bone marrow stromal cells affords a significant level of protection against chemotherapy. Thus, it is conceivable that inhibiting CTGF function using blocking antibodies during induction chemotherapy could hypothetically reduce the emergence of resistance. There is much to be learnt from murine models before such a hypothesis could be tested, however attenuating CTGF function could conceivably provide a novel therapeutic strategy for the management of paediatric pre-B ALL in the future.

7.4.3 Factors interacting with the *CTGF* gene locus

While nuclear factors known to regulate the *CTGF* locus were not expressed at higher levels in *CTGF*^{POS} cells, these factors may be activated post-transcriptionally, or through aberrant activation of binding partners. As such, analysis of nuclear factors bound at the *CTGF* gene locus should be explored by chromatin immunoprecipitation (ChIP). Candidates for analysis by ChIP include those factors known to activate the *CTGF* promoter in other cell types such as AP-1, Ets-1, MMP3 or SP-1 (Yu et al. 2009; Van Beek et al. 2006; Eguchi et al. 2008; Holmes et al. 2003), as well as members of the Ikaros family of nuclear factors highlighted in the present study.

7.5 Final conclusions

This thesis supports previous observations that *CTGF* expression is associated with a poor outcome in paediatric pre-B ALL and identified an association between *CTGF*

gene expression and reduced lymph node metastasis, suggesting *CTGF* expression may influence homing of leukaemic cells. A disparity between *CTGF* mRNA and protein levels was observed, thus future studies should develop methods for detecting *CTGF* protein levels in the bone marrow space. Analysis of global gene expression data from three independent cohorts implicated the Ikaros family lymphoid-specific nuclear factors in regulation of the *CTGF* gene locus. Functional investigations into the role of these lymphoid-specific nuclear factors in altering *CTGF* expression are now warranted. Genomic lesions such as gene rearrangements, copy number alterations, and gene mutations were ruled out as mechanisms contributing to *CTGF* expression in pre-B ALL, however hypomethylation of the *CTGF* locus was a feature of all primary pre-B ALL specimens examined, providing a plausible mechanism for its deregulated expression. As well as aberrant epigenetic marks, alternative splicing of *CTGF* mRNA was observed and minor transcripts were characterized that displayed altered coding potential and others with the ability to evade post-transcriptional control mechanisms.

In conclusions this thesis has improved our understanding of the molecular mechanisms contributing to *CTGF* gene expression in pre-B ALL, as well as highlighting important aspects of *CTGF* biology that require further investigation. Continued research into the role of *CTGF* gene expression in the development and maintenance of pre-B ALL may uncover biological interactions that are critical for ALL pathogenesis. There is evidence emerging from both *in vitro* and *in vivo* models to suggest that inappropriate *CTGF* secretion by leukaemic cells may have significant effects upon the establishment and progression of this disease. Furthermore these interactions may be therapeutically vulnerable and may represent a valid avenue of improving patient therapy in the future. By continuing to advance our understanding of key molecular pathways underlying ALL, we can move toward improving treatments and ultimately long-term outcomes for children diagnosed with this disease.

Chapter 8

References

References

Declaration

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Abreu, J. G., Ketpura, N. I., Reversade, B. & De Robertis, E. M. 2002, 'Connective-tissue growth factor (CTGF) modulates cell signalling by BMP and TGF-beta', *Nat Cell Biol*, vol. 4, no. 8, pp. 599-604.

Agueli, C., Cammarata, G., Salemi, D., Dagnino, L., Nicoletti, R., La Rosa, M., Messina, F., Marfia, A., Bica, M. G., Coniglio, M. L., Pagano, M., Fabbiano, F. & Santoro, A. 2010, '14q32/miRNA clusters loss of heterozygosity in acute lymphoblastic leukemia is associated with up-regulation of BCL11a', *Am J Hematol*, vol. 85, no. 8, pp. 575-8.

Aikawa, T., Gunn, J., Spong, S. M., Klaus, S. J. & Korc, M. 2006, 'Connective tissue growth factor-specific antibody attenuates tumor growth, metastasis, and angiogenesis in an orthotopic mouse model of pancreatic cancer', *Mol Cancer Ther*, vol. 5, no. 5, pp. 1108-16.

Akiyama, Y. n.d., *TFSEARCH: Searching transcription factor binding sites*.

Albihn, A., Johnsen, J. I. & Henriksson, M. A. 2010, 'MYC in oncogenesis and as a target for cancer therapies', *Adv Cancer Res*, vol. 107, pp. 163-224.

Allen, M. & Louise Jones, J. 2011, 'Jekyll and Hyde: the role of the microenvironment on the progression of cancer', *J Pathol*, vol. 223, no. 2, pp. 162-76.

Amati, B., Frank, S. R., Donjerkovic, D. & Taubert, S. 2001, 'Function of the c-Myc oncoprotein in chromatin remodeling and transcription', *Biochim Biophys Acta*, vol. 1471, no. 3, pp. M135-45.

Annino, L., Vegna, M. L., Camera, A., Specchia, G., Visani, G., Fioritoni, G., Ferrara, F., Peta, A., Ciolli, S., Deplano, W., Fabbiano, F., Sica, S., Di Raimondo, F., Cascavilla, N., Tabilio, A., Leoni, P., Invernizzi, R., Baccarani, M., Rotoli, B., Amadori, S. & Mandelli, F. 2002, 'Treatment of adult acute lymphoblastic leukemia (ALL): long-term follow-up of the GIMEMA ALL 0288 randomized study', *Blood*, vol. 99, no. 3, pp. 863-71.

Arico, M., Valsecchi, M. G., Camitta, B., Schrappe, M., Chessells, J., Baruchel, A., Gaynon, P., Silverman, L., Janka-Schaub, G., Kamps, W., Pui, C. H. & Masera, G.

2000, 'Outcome of treatment in children with Philadelphia chromosome-positive acute lymphoblastic leukemia', *N Engl J Med*, vol. 342, no. 14, pp. 998-1006.

Arico, M., Valsecchi, M. G., Rizzari, C., Barisone, E., Biondi, A., Casale, F., Locatelli, F., Lo Nigro, L., Luciani, M., Messina, C., Micalizzi, C., Parasole, R., Pession, A., Santoro, N., Testi, A. M., Silvestri, D., Basso, G., Masera, G. & Conter, V. 2008, 'Long-term results of the AIEOP-ALL-95 Trial for Childhood Acute Lymphoblastic Leukemia: insight on the prognostic value of DNA index in the framework of Berlin-Frankfurt-Muenster based chemotherapy', *J Clin Oncol*, vol. 26, no. 2, pp. 283-9.

Arnott, J. A., Zhang, X., Sanjay, A., Owen, T. A., Smock, S. L., Rehman, S., DeLong, W. G., Safadi, F. F. & Popoff, S. N. 2008, 'Molecular requirements for induction of CTGF expression by TGF-beta1 in primary osteoblasts', *Bone*, vol. 42, no. 5, pp. 871-85.

Aspland, S. E., Bendall, H. H. & Murre, C. 2001, 'The role of E2A-PBX1 in leukemogenesis', *Oncogene*, vol. 20, no. 40, pp. 5708-17.

Australian Bureau of Statistics 2006, *Causes of Death, Australia (2004)*, 3303.0, Australian Bureau of Statistics, Canberra.

Ayton, P. M. & Cleary, M. L. 2001, 'Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins', *Oncogene*, vol. 20, no. 40, pp. 5695-707.

Bailey, T. L. & Elkan, C. 1994, 'Fitting a mixture model by expectation maximization to discover motifs in biopolymers', *Proc Int Conf Intell Syst Mol Biol*, vol. 2, pp. 28-36.

Bain, G., Romanow, W. J., Albers, K., Havran, W. L. & Murre, C. 1999, 'Positive and negative regulation of V(D)J recombination by the E2A proteins', *J Exp Med*, vol. 189, no. 2, pp. 289-300.

Ball, D. K., Moussad, E. E., Rageh, M. A., Kemper, S. A. & Brigstock, D. R. 2003a, 'Establishment of a recombinant expression system for connective tissue growth factor (CTGF) that models CTGF processing in utero', *Reproduction*, vol. 125, no. 2, pp. 271-84.

Ball, D. K., Rachfal, A. W., Kemper, S. A. & Brigstock, D. R. 2003b, 'The heparin-binding 10 kDa fragment of connective tissue growth factor (CTGF) containing module 4 alone stimulates cell adhesion', *J Endocrinol*, vol. 176, no. 2, pp. R1-7.

Ball, D. K., Surveyor, G. A., Diehl, J. R., Steffen, C. L., Uzumcu, M., Mirando, M. A. & Brigstock, D. R. 1998, 'Characterization of 16- to 20-kilodalton (kDa) connective tissue growth factors (CTGFs) and demonstration of proteolytic activity for 38-kDa CTGF in pig uterine luminal flushings', *Biol Reprod*, vol. 59, no. 4, pp. 828-35.

Barbolina, M. V., Adley, B. P., Kelly, D. L., Shepard, J., Fought, A. J., Scholtens, D., Penzes, P., Shea, L. D. & Stack, M. S. 2009, 'Downregulation of connective tissue growth factor by three-dimensional matrix enhances ovarian carcinoma cell invasion', *Int J Cancer*, vol. 125, no. 4, pp. 816-25.

Barth, P. J. & Westhoff, C. C. 2007, 'CD34+ fibrocytes: morphology, histogenesis and function', *Curr Stem Cell Res Ther*, vol. 2, no. 3, pp. 221-7.

Beesley, A. H., Palmer, M. L., Ford, J., Weller, R. E., Cummings, A. J., Freitas, J. R., Firth, M. J., Perera, K. U., de Klerk, N. H. & Kees, U. R. 2006, 'Authenticity and drug resistance in a panel of acute lymphoblastic leukaemia cell lines', *Br J Cancer*, vol. 95, no. 11, pp. 1537-44.

Bene, M. C. 2005, 'Immunophenotyping of acute leukaemias', *Immunol Lett*, vol. 98, no. 1, pp. 9-21.

Bernstein, B. E., Meissner, A. & Lander, E. S. 2007, 'The mammalian epigenome', *Cell*, vol. 128, no. 4, pp. 669-81.

Bhatia, S. 2004, 'Influence of race and socioeconomic status on outcome of children treated for childhood acute lymphoblastic leukemia', *Curr Opin Pediatr*, vol. 16, no. 1, pp. 9-14.

Bjourson, A. J. & Cooper, J. E. 1992, 'Band-stab PCR: a simple technique for the purification of individual PCR products', *Nucleic Acids Res*, vol. 20, no. 17, p. 4675.

Blackledge, N. P., Zhou, J. C., Tolstorukov, M. Y., Farcas, A. M., Park, P. J. & Klose, R. J. 2010, 'CpG islands recruit a histone H3 lysine 36 demethylase', *Mol Cell*, vol. 38, no. 2, pp. 179-90.

Blesa, S., Vernia, S., Garcia-Garcia, A. B., Martinez-Hervas, S., Ivorra, C., Gonzalez-Albert, V., Ascaso, J. F., Martin-Escudero, J. C., Real, J. T., Carmena, R., Casado, M. & Chaves, F. J. 2008, 'A new PCSK9 gene promoter variant affects gene expression and causes autosomal dominant hypercholesterolemia', *J Clin Endocrinol Metab*, vol. 93, no. 9, pp. 3577-83.

Blom, B. & Spits, H. 2006, 'Development of human lymphoid cells', *Annu Rev Immunol*, vol. 24, pp. 287-320.

Boag, J. M., Beesley, A. H., Firth, M. J., Freitas, J. R., Ford, J., Brigstock, D. R., de Klerk, N. H. & Kees, U. R. 2007, 'High expression of connective tissue growth factor in pre-B acute lymphoblastic leukaemia', *Br J Haematol*, vol. 138, no. 6, pp. 740-8.

Boag, J. M., Beesley, A. H., Firth, M. J., Freitas, J. R., Ford, J., Hoffmann, K., Cummings, A. J., de Klerk, N. H. & Kees, U. R. 2006, 'Altered glucose metabolism in childhood pre-B acute lymphoblastic leukaemia', *Leukemia*, vol. 20, no. 10, pp. 1731-7.

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J. & Lengauer, T. 2005, 'BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing', *Bioinformatics*, vol. 21, no. 21, pp. 4067-8.

Boidot, R., Vegran, F., Jacob, D., Chevrier, S., Cadouot, M., Feron, O., Solary, E. & Lizard-Nacol, S. 2010, 'The transcription factor GATA-1 is overexpressed in breast carcinomas and contributes to survivin upregulation via a promoter polymorphism', *Oncogene*, vol. 29, no. 17, pp. 2577-84.

Bonifer, C. & Bowen, D. T. 2010, 'Epigenetic mechanisms regulating normal and malignant haematopoiesis: new therapeutic targets for clinical medicine', *Expert Rev Mol Med*, vol. 12, p. e6.

Bourquin, J. P. & Izraeli, S. 2010, 'Where can biology of childhood ALL be attacked by new compounds?' *Cancer Treat Rev*, vol. 36, no. 4, pp. 298-306.

Bradford, M. M. 1976, 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Anal Biochem*, vol. 72, pp. 248-54.

Bradham, D. M., Igarashi, A., Potter, R. L. & Grotendorst, G. R. 1991, 'Connective tissue growth factor: a cysteine-rich mitogen secreted by human vascular endothelial cells is related to the SRC-induced immediate early gene product CEF-10', *J Cell Biol*, vol. 114, no. 6, pp. 1285-94.

Brigstock, D. R. 2003, 'The CCN family: a new stimulus package', *J Endocrinol*, vol. 178, no. 2, pp. 169-75.

Brigstock, D. R. 2010, 'Connective tissue growth factor (CCN2, CTGF) and organ fibrosis: lessons from transgenic animals', *J Cell Commun Signal*, vol. 4, no. 1, pp. 1-4.

Brigstock, D. R., Steffen, C. L., Kim, G. Y., Vegunta, R. K., Diehl, J. R. & Harding, P. A. 1997, 'Purification and characterization of novel heparin-binding growth factors in uterine secretory fluids. Identification as heparin-regulated Mr 10,000 forms of connective tissue growth factor', *J Biol Chem*, vol. 272, no. 32, pp. 20275-82.

Brothman, A. R., Swanson, G., Maxwell, T. M., Cui, J., Murphy, K. J., Herrick, J., Speights, V. O., Isaac, J. & Rohr, L. R. 2005, 'Global hypomethylation is common in prostate cancer cells: a quantitative predictor for clinical outcome?' *Cancer Genet Cytogenet*, vol. 156, no. 1, pp. 31-6.

Brown, P., Levis, M., Shurtleff, S., Campana, D., Downing, J. & Small, D. 2005, 'FLT3 inhibition selectively kills childhood acute lymphoblastic leukemia cells with high levels of FLT3 expression', *Blood*, vol. 105, no. 2, pp. 812-20.

Bruggemann, M., Schrauder, A., Raff, T., Pfeifer, H., Dworzak, M., Ottmann, O. G., Asnafi, V., Baruchel, A., Bassan, R., Benoit, Y., Biondi, A., Cave, H., Dombret, H., Fielding, A. K., Foa, R., Gokbuget, N., Goldstone, A. H., Goulden, N., Henze, G., Hoelzer, D., Janka-Schaub, G. E., Macintyre, E. A., Pieters, R., Rambaldi, A., Ribera, J. M., Schmiegelow, K., Spinelli, O., Stary, J., von Stackelberg, A., Kneba, M., Schrappe, M. & van Dongen, J. J. 2010, 'Standardized MRD quantification in European ALL trials: proceedings of the Second International Symposium on MRD assessment in Kiel, Germany, 18-20 September 2008', *Leukemia*, vol. 24, no. 3, pp. 521-35.

Buffler, P. A., Kwan, M. L., Reynolds, P. & Urayama, K. Y. 2005, 'Environmental and genetic risk factors for childhood leukemia: appraising the evidence', *Cancer Invest*, vol. 23, no. 1, pp. 60-75.

Cadioux, B., Ching, T. T., VandenBerg, S. R. & Costello, J. F. 2006, 'Genome-wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation', *Cancer Res*, vol. 66, no. 17, pp. 8469-76.

Cervello, M., Giannitrapani, L., Labbozzetta, M., Notarbartolo, M., D'Alessandro, N., Lampiasi, N., Azzolina, A. & Montalto, G. 2004, 'Expression of WISPs and of their novel alternative variants in human hepatocellular carcinoma cells', *Ann N Y Acad Sci*, vol. 1028, pp. 432-9.

Chaqour, B. & Goppelt-Struebe, M. 2006, 'Mechanical regulation of the Cyr61/CCN1 and CTGF/CCN2 proteins', *FEBS J*, vol. 273, no. 16, pp. 3639-49.

Chen, C. C. & Lau, L. F. 2009, 'Functions and mechanisms of action of CCN matricellular proteins', *Int J Biochem Cell Biol*, vol. 41, no. 4, pp. 771-83.

Chen, N., Chen, C. C. & Lau, L. F. 2000, 'Adhesion of human skin fibroblasts to Cyr61 is mediated through integrin alpha 6beta 1 and cell surface heparan sulfate proteoglycans', *J Biol Chem*, vol. 275, no. 32, pp. 24953-61.

Chen, Y., Blom, I. E., Sa, S., Goldschmeding, R., Abraham, D. J. & Leask, A. 2002, 'CTGF expression in mesangial cells: involvement of SMADs, MAP kinase, and PKC', *Kidney Int*, vol. 62, no. 4, pp. 1149-59.

Chen, Y., Segarini, P., Raoufi, F., Bradham, D. & Leask, A. 2001, 'Connective tissue growth factor is secreted through the Golgi and is degraded in the endosome', *Exp Cell Res*, vol. 271, no. 1, pp. 109-17.

Chi, P., Allis, C. D. & Wang, G. G. 2010, 'Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers', *Nat Rev Cancer*, vol. 10, no. 7, pp. 457-69.

Chiba, T., Yokosuka, O., Arai, M., Tada, M., Fukai, K., Imazeki, F., Kato, M., Seki, N. & Saisho, H. 2004, 'Identification of genes up-regulated by histone deacetylase

inhibition with cDNA microarray and exploration of epigenetic alterations on hepatoma cells', *J Hepatol*, vol. 41, no. 3, pp. 436-45.

Chiba, T., Yokosuka, O., Fukai, K., Hirasawa, Y., Tada, M., Mikata, R., Imazeki, F., Taniguchi, H., Iwama, A., Miyazaki, M., Ochiai, T. & Saisho, H. 2005, 'Identification and investigation of methylated genes in hepatoma', *Eur J Cancer*, vol. 41, no. 8, pp. 1185-94.

Chien, W., Yin, D., Gui, D., Mori, A., Frank, J. M., Said, J., Kusuanco, D., Marchevsky, A., McKenna, R. & Koeffler, H. P. 2006, 'Suppression of cell proliferation and signaling transduction by connective tissue growth factor in non-small cell lung cancer cells', *Mol Cancer Res*, vol. 4, no. 8, pp. 591-8.

Choi, J., Hwang, Y. K., Sung, K. W., Kim, D. H., Yoo, K. H., Jung, H. L. & Koo, H. H. 2006, 'Aven overexpression: association with poor prognosis in childhood acute lymphoblastic leukemia', *Leuk Res*, vol. 30, no. 8, pp. 1019-25.

Chouaib, S., Kieda, C., Benlalam, H., Noman, M. Z., Mami-Chouaib, F. & Ruegg, C. 2011, 'Endothelial cells as key determinants of the tumor microenvironment: interaction with tumor cells, extracellular matrix and immune killer cells', *Crit Rev Immunol*, vol. 30, no. 6, pp. 529-45.

Chu, C. Y., Chang, C. C., Prakash, E. & Kuo, M. L. 2008, 'Connective tissue growth factor (CTGF) and cancer progression', *J Biomed Sci*, vol. 15, no. 6, pp. 675-85.

Cicha, I. & Goppelt-Struebe, M. 2009, 'Connective tissue growth factor: context-dependent functions and mechanisms of regulation', *Biofactors*, vol. 35, no. 2, pp. 200-8.

Cobb, B. S. & Smale, S. T. 2005, 'Ikaros-family proteins: in search of molecular functions during lymphocyte development', *Curr Top Microbiol Immunol*, vol. 290, pp. 29-47.

Conter, V., Arico, M., Basso, G., Biondi, A., Barisone, E., Messina, C., Parasole, R., De Rossi, G., Locatelli, F., Pession, A., Santoro, N., Micalizzi, C., Citterio, M., Rizzari, C., Silvestri, D., Rondelli, R., Lo Nigro, L., Ziino, O., Testi, A. M., Masera, G. & Valsecchi, M. G. 2010, 'Long-term results of the Italian Association of Pediatric Hematology and Oncology (AIEOP) Studies 82, 87, 88, 91 and 95 for childhood acute lymphoblastic leukemia', *Leukemia*, vol. 24, no. 2, pp. 255-64.

Corfe, S. & Paige, C. 2009, 'Development of B Lymphocytes', in *Molecular Basis of Hematopoiesis*, eds A. Wickrema & B. Kee, Springer, New York, pp. 173-199.

Crisan, M., Yap, S., Casteilla, L., Chen, C. W., Corselli, M., Park, T. S., Andriolo, G., Sun, B., Zheng, B., Zhang, L., Norotte, C., Teng, P. N., Traas, J., Schugar, R., Deasy, B. M., Badylak, S., Buhning, H. J., Giacobino, J. P., Lazzari, L., Huard, J. &

Peault, B. 2008, 'A perivascular origin for mesenchymal stem cells in multiple human organs', *Cell Stem Cell*, vol. 3, no. 3, pp. 301-13.

De Angioletti, M., Lacerra, G., Sabato, V. & Carestia, C. 2004, 'Beta+45 G --> C: a novel silent beta-thalassaemia mutation, the first in the Kozak sequence', *Br J Haematol*, vol. 124, no. 2, pp. 224-31.

de Capoa, A., Musolino, A., Della Rosa, S., Caiafa, P., Mariani, L., Del Nonno, F., Vocaturo, A., Donnorso, R. P., Niveleau, A. & Grappelli, C. 2003, 'DNA demethylation is directly related to tumour progression: evidence in normal, pre-malignant and malignant cells from uterine cervix samples', *Oncol Rep*, vol. 10, no. 3, pp. 545-9.

Denoeud, F., Kapranov, P., Ucla, C., Frankish, A., Castelo, R., Drenkow, J., Lagarde, J., Alioto, T., Manzano, C., Chrast, J., Dike, S., Wyss, C., Henrichsen, C. N., Holroyd, N., Dickson, M. C., Taylor, R., Hance, Z., Foissac, S., Myers, R. M., Rogers, J., Hubbard, T., Harrow, J., Guigo, R., Gingeras, T. R., Antonarakis, S. E. & Reymond, A. 2007, 'Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions', *Genome Res*, vol. 17, no. 6, pp. 746-59.

Dhar, A. & Ray, A. 2010, 'The CCN family proteins in carcinogenesis', *Exp Oncol*, vol. 32, no. 1, pp. 2-9.

Dias, S., Silva, H., Jr., Cumano, A. & Vieira, P. 2005, 'Interleukin-7 is necessary to maintain the B cell potential in common lymphoid progenitors', *J Exp Med*, vol. 201, no. 6, pp. 971-9.

Dowling, E., Yabroff, K. R., Mariotto, A., McNeel, T., Zeruto, C. & Buckman, D. 2010, 'Burden of illness in adult survivors of childhood cancers: findings from a population-based national sample', *Cancer*, vol. 116, no. 15, pp. 3712-21.

Duisters, R. F., Tijssen, A. J., Schroen, B., Leenders, J. J., Lentink, V., van der Made, I., Herias, V., van Leeuwen, R. E., Schellings, M. W., Barenbrug, P., Maessen, J. G., Heymans, S., Pinto, Y. M. & Creemers, E. E. 2009, 'miR-133 and miR-30 regulate connective tissue growth factor: implications for a role of microRNAs in myocardial matrix remodeling', *Circ Res*, vol. 104, no. 2, pp. 170-8.

Ebinger, M., Senf, L., Wachowski, O. & Scheurlen, W. 2004, 'Promoter methylation pattern of caspase-8, P16INK4A, MGMT, TIMP-3, and E-cadherin in medulloblastoma', *Pathol Oncol Res*, vol. 10, no. 1, pp. 17-21.

Eguchi, T., Kubota, S., Kawata, K., Mukudai, Y., Ohgawara, T., Miyazono, K., Nakao, K., Kondo, S. & Takigawa, M. 2007, 'Different transcriptional strategies for *ccn2/ctgf* gene induction between human chondrocytic and breast cancer cell lines', *Biochimie*, vol. 89, no. 3, pp. 278-88.

- Eguchi, T., Kubota, S., Kawata, K., Mukudai, Y., Uehara, J., Ohgawara, T., Ibaragi, S., Sasaki, A., Kuboki, T. & Takigawa, M. 2008, 'Novel transcription-factor-like function of human matrix metalloproteinase 3 regulating the CTGF/CCN2 gene', *Mol Cell Biol*, vol. 28, no. 7, pp. 2391-413.
- Ehninger, A. & Trumpp, A. 2011, 'The bone marrow stem cell niche grows up: mesenchymal stem cells and macrophages move in', *J Exp Med*, vol. 208, no. 3, pp. 421-8.
- Ehrlich, M. 2009, 'DNA hypomethylation in cancer cells', *Epigenomics*, vol. 1, no. 2, pp. 239-259.
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. 1998, 'Cluster analysis and display of genome-wide expression patterns', *Proc Natl Acad Sci U S A*, vol. 95, no. 25, pp. 14863-8.
- Elmlinger, M. W., Wimmer, K., Biemer, E., Blum, W. F., Ranke, M. B. & Dannecker, G. E. 1996, 'Insulin-like growth factor binding protein 2 is differentially expressed in leukaemic B- and T-cell lines', *Growth Regul*, vol. 6, no. 3, pp. 152-7.
- Ernst, A., Campos, B., Meier, J., Devens, F., Liesenberg, F., Wolter, M., Reifemberger, G., Herold-Mende, C., Lichter, P. & Radlwimmer, B. 2010, 'De-repression of CTGF via the miR-17-92 cluster upon differentiation of human glioblastoma spheroid cultures', *Oncogene*, vol. 29, no. 23, pp. 3411-22.
- Farre, D., Roset, R., Huerta, M., Adsuara, J. E., Rosello, L., Alba, M. M. & Messeguer, X. 2003, 'Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN', *Nucleic Acids Res*, vol. 31, no. 13, pp. 3651-3.
- Faull, R. J. 2005, 'Bad and good growth factors in the peritoneal cavity', *Nephrology*, vol. 10, no. 3, pp. 234-9.
- Feinberg, A. P. & Vogelstein, B. 1983a, 'Hypomethylation distinguishes genes of some human cancers from their normal counterparts', *Nature*, vol. 301, no. 5895, pp. 89-92.
- Feinberg, A. P. & Vogelstein, B. 1983b, 'Hypomethylation of ras oncogenes in primary human cancers', *Biochem Biophys Res Commun*, vol. 111, no. 1, pp. 47-54.
- Feng, Q. & Zhang, Y. 2001, 'The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes', *Genes Dev*, vol. 15, no. 7, pp. 827-32.
- Fischer, M., Schwieger, M., Horn, S., Niebuhr, B., Ford, A., Roscher, S., Bergholz, U., Greaves, M., Lohler, J. & Stocking, C. 2005, 'Defining the oncogenic function of

the TEL/AML1 (ETV6/RUNX1) fusion protein in a mouse model', *Oncogene*, vol. 24, no. 51, pp. 7579-91.

Fonseca, C., Lindahl, G. E., Ponticos, M., Sestini, P., Renzoni, E. A., Holmes, A. M., Spagnolo, P., Pantelidis, P., Leoni, P., McHugh, N., Stock, C. J., Shi-Wen, X., Denton, C. P., Black, C. M., Welsh, K. I., du Bois, R. M. & Abraham, D. J. 2007, 'A polymorphism in the CTGF promoter region associated with systemic sclerosis', *N Engl J Med*, vol. 357, no. 12, pp. 1210-20.

Frazier, K. S. & Grotendorst, G. R. 1997, 'Expression of connective tissue growth factor mRNA in the fibrous stroma of mammary tumors', *Int J Biochem Cell Biol*, vol. 29, no. 1, pp. 153-61.

Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. 2009, 'Most mammalian mRNAs are conserved targets of microRNAs', *Genome Res*, vol. 19, no. 1, pp. 92-105.

Fu, M., Zhang, J., Zhu, X., Myles, D. E., Willson, T. M., Liu, X. & Chen, Y. E. 2001, 'Peroxisome proliferator-activated receptor gamma inhibits transforming growth factor beta-induced connective tissue growth factor expression in human aortic smooth muscle cells by interfering with Smad3', *J Biol Chem*, vol. 276, no. 49, pp. 45888-94.

Fu, Y. & Weng, Z. 2004, 'Improvement of TRANSFAC matrices using multiple local alignment of transcription factor binding site sequences', *Conf Proc IEEE Eng Med Biol Soc*, vol. 4, pp. 2856-9.

Gala, J. L., Vermynen, C., Cornu, G., Ferrant, A., Michaux, J. L., Philippe, M. & Martiat, P. 1994, 'High expression of bcl-2 is the rule in acute lymphoblastic leukemia, except in Burkitt subtype at presentation, and is not correlated with the prognosis', *Ann Hematol*, vol. 69, no. 1, pp. 17-24.

Gama-Sosa, M. A., Slagel, V. A., Trewyn, R. W., Oxenhandler, R., Kuo, K. C., Gehrke, C. W. & Ehrlich, M. 1983, 'The 5-methylcytosine content of DNA from human tumors', *Nucleic Acids Res*, vol. 11, no. 19, pp. 6883-94.

Gao, R. & Brigstock, D. R. 2003, 'Low density lipoprotein receptor-related protein (LRP) is a heparin-dependent adhesion receptor for connective tissue growth factor (CTGF) in rat activated hepatic stellate cells', *Hepatol Res*, vol. 27, no. 3, pp. 214-220.

Gao, R. & Brigstock, D. R. 2004, 'Connective tissue growth factor (CCN2) induces adhesion of rat activated hepatic stellate cells by binding of its C-terminal domain to integrin alpha(v)beta(3) and heparan sulfate proteoglycan', *J Biol Chem*, vol. 279, no. 10, pp. 8848-55.

Gao, R. & Brigstock, D. R. 2006, 'A novel integrin alpha5beta1 binding domain in module 4 of connective tissue growth factor (CCN2/CTGF) promotes adhesion and migration of activated pancreatic stellate cells', *Gut*, vol. 55, no. 6, pp. 856-62.

Gaynon, P. S. 2005, 'Childhood acute lymphoblastic leukaemia and relapse', *Br J Haematol*, vol. 131, no. 5, pp. 579-87.

Gaynon, P. S., Angiolillo, A. L., Carroll, W. L., Nachman, J. B., Trigg, M. E., Sather, H. N., Hunger, S. P. & Devidas, M. 2010, 'Long-term results of the children's cancer group studies for childhood acute lymphoblastic leukemia 1983-2002: a Children's Oncology Group Report', *Leukemia*, vol. 24, no. 2, pp. 285-97.

Gaynon, P. S., Trigg, M. E., Heerema, N. A., Sensel, M. G., Sather, H. N., Hammond, G. D. & Bleyer, W. A. 2000, 'Children's Cancer Group trials in childhood acute lymphoblastic leukemia: 1983-1995', *Leukemia*, vol. 14, no. 12, pp. 2223-33.

Geenen, M. M., Bakker, P. J., Kremer, L. C., Kastelein, J. J. & van Leeuwen, F. E. 2010, 'Increased prevalence of risk factors for cardiovascular disease in long-term survivors of acute lymphoblastic leukemia and Wilms tumor treated with radiotherapy', *Pediatr Blood Cancer*, vol. 55, no. 4, pp. 690-7.

Georgopoulos, K., Bigby, M., Wang, J. H., Molnar, A., Wu, P., Winandy, S. & Sharpe, A. 1994, 'The Ikaros gene is required for the development of all lymphoid lineages', *Cell*, vol. 79, no. 1, pp. 143-56.

Gourh, P., Mayes, M. D. & Arnett, F. C. 2008, 'CTGF polymorphism associated with systemic sclerosis', *N Engl J Med*, vol. 358, no. 3, pp. 308-9; author reply 309.

Granel, B., Argiro, L., Hachulla, E., Fajardy, I., Weiller, P. J., Durand, J. M., Frances, Y., Dombey, A. M., Marquet, S., Lesavre, N., Disdier, P., Bernard, F., Hatron, P. Y. & Chevillard, C. 2010, 'Association between a CTGF gene polymorphism and systemic sclerosis in a French population', *J Rheumatol*, vol. 37, no. 2, pp. 351-8.

Greaves, M. 2009, 'Darwin and evolutionary tales in leukemia. The Ham-Wasserman Lecture', *Hematology Am Soc Hematol Educ Program*, pp. 3-12.

Gressner, O. A. & Gressner, A. M. 2008, 'Connective tissue growth factor: a fibrogenic master switch in fibrotic liver diseases', *Liver Int*, vol. 28, no. 8, pp. 1065-79.

Grotendorst, G. R. 1997, 'Connective tissue growth factor: a mediator of TGF-beta action on fibroblasts', *Cytokine Growth Factor Rev*, vol. 8, no. 3, pp. 171-9.

Grotendorst, G. R., Okochi, H. & Hayashi, N. 1996, 'A novel transforming growth factor beta response element controls the expression of the connective tissue growth factor gene', *Cell Growth Differ*, vol. 7, no. 4, pp. 469-80.

Guenther, M. G., Lawton, L. N., Rozovskaia, T., Frampton, G. M., Levine, S. S., Volkert, T. L., Croce, C. M., Nakamura, T., Canaani, E. & Young, R. A. 2008, 'Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia', *Genes Dev*, vol. 22, no. 24, pp. 3403-8.

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. 2007, 'Quantifying similarity between motifs', *Genome Biol*, vol. 8, no. 2, p. R24.

Gurney, J. G., Severson, R. K., Davis, S. & Robison, L. L. 1995, 'Incidence of cancer in children in the United States. Sex-, race-, and 1-year age-specific rates by histologic type', *Cancer*, vol. 75, no. 8, pp. 2186-95.

Hallegger, M., Llorian, M. & Smith, C. W. 2010, 'Alternative splicing: global insights', *FEBS J*, vol. 277, no. 4, pp. 856-66.

Hanada, M., Delia, D., Aiello, A., Stadtmauer, E. & Reed, J. C. 1993, 'bcl-2 gene hypomethylation and high-level expression in B-cell chronic lymphocytic leukemia', *Blood*, vol. 82, no. 6, pp. 1820-8.

Harrison, C. J., Moorman, A. V., Broadfield, Z. J., Cheung, K. L., Harris, R. L., Reza Jalali, G., Robinson, H. M., Barber, K. E., Richards, S. M., Mitchell, C. D., Eden, T. O., Hann, I. M., Hill, F. G., Kinsey, S. E., Gibson, B. E., Lilleyman, J., Vora, A., Goldstone, A. H., Franklin, I. M., Durrant, J. & Martineau, M. 2004, 'Three distinct subgroups of hypodiploidy in acute lymphoblastic leukaemia', *Br J Haematol*, vol. 125, no. 5, pp. 552-9.

Harvey, R. C., Mullighan, C. G., Wang, X., Dobbin, K. K., Davidson, G. S., Bedrick, E. J., Chen, I. M., Atlas, S. R., Kang, H., Ar, K., Wilson, C. S., Wharton, W., Murphy, M., Devidas, M., Carroll, A. J., Borowitz, M. J., Bowman, W. P., Downing, J. R., Relling, M., Yang, J., Bhojwani, D., Carroll, W. L., Camitta, B., Reaman, G. H., Smith, M., Hunger, S. P. & Willman, C. L. 2010, 'Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome', *Blood*, vol. 116, no. 23, pp. 4874-84.

Hashimoto, G., Inoki, I., Fujii, Y., Aoki, T., Ikeda, E. & Okada, Y. 2002, 'Matrix metalloproteinases cleave connective tissue growth factor and reactivate angiogenic activity of vascular endothelial growth factor 165', *J Biol Chem*, vol. 277, no. 39, pp. 36288-95.

Hatakeyama, M., Tsudo, M., Minamoto, S., Kono, T., Doi, T., Miyata, T., Miyasaka, M. & Taniguchi, T. 1989, 'Interleukin-2 receptor beta chain gene: generation of three receptor forms by cloned human alpha and beta chain cDNA's', *Science*, vol. 244, no. 4904, pp. 551-6.

He, S., Jin, M. L., Worpel, V. & Hinton, D. R. 2003, 'A role for connective tissue growth factor in the pathogenesis of choroidal neovascularization', *Arch Ophthalmol*, vol. 121, no. 9, pp. 1283-8.

Heerema, N. A., Sather, H. N., Sensel, M. G., Zhang, T., Hutchinson, R. J., Nachman, J. B., Lange, B. J., Steinherz, P. G., Bostrom, B. C., Reaman, G. H., Gaynon, P. S. & Uckun, F. M. 2000, 'Prognostic impact of trisomies of chromosomes 10, 17, and 5 among children with acute lymphoblastic leukemia and high hyperdiploidy (> 50 chromosomes)', *J Clin Oncol*, vol. 18, no. 9, pp. 1876-87.

Higgins, D. F., Biju, M. P., Akai, Y., Wutz, A., Johnson, R. S. & Haase, V. H. 2004, 'Hypoxic induction of Ctgf is directly mediated by Hif-1', *Am J Physiol Renal Physiol*, vol. 287, no. 6, pp. F1223-32.

Hirschfeld, M., zur Hausen, A., Bettendorf, H., Jager, M. & Stickeler, E. 2009, 'Alternative splicing of Cyr61 is regulated by hypoxia and significantly changed in breast cancer', *Cancer Res*, vol. 69, no. 5, pp. 2082-90.

Hjalgrim, L. L., Madsen, H. O., Melbye, M., Jorgensen, P., Christiansen, M., Andersen, M. T., Pallisgaard, N., Hokland, P., Clausen, N., Ryder, L. P., Schmiegelow, K. & Hjalgrim, H. 2002, 'Presence of clone-specific markers at birth in children with acute lymphoblastic leukaemia', *Br J Cancer*, vol. 87, no. 9, pp. 994-9.

Ho, J. W., Bishop, E., Kharchenko, P. V., Negre, N., White, K. P. & Park, P. J. 2011, 'ChIP-chip versus ChIP-seq: Lessons for experimental design and data analysis', *BMC Genomics*, vol. 12, no. 1, p. 134.

Hoffmann, K., Firth, M. J., Beesley, A. H., de Klerk, N. H. & Kees, U. R. 2006, 'Translating microarray data for diagnostic testing in childhood leukaemia', *BMC Cancer*, vol. 6, p. 229-240.

Hoffmann, K., Firth, M. J., Freitas, J. R., de Klerk, N. H. & Kees, U. R. 2005, 'Gene expression levels in small specimens from patients detected using oligonucleotide arrays', *Mol Biotechnol*, vol. 29, no. 1, pp. 31-8.

Holbourn, K. P., Acharya, K. R. & Perbal, B. 2008, 'The CCN family of proteins: structure-function relationships', *Trends Biochem Sci*, vol. 33, no. 10, pp. 461-73.

Holliday, R. & Pugh, J. E. 1975, 'DNA modification mechanisms and gene activity during development', *Science*, vol. 187, no. 4173, pp. 226-32.

Holmes, A., Abraham, D. J., Chen, Y., Denton, C., Shi-wen, X., Black, C. M. & Leask, A. 2003, 'Constitutive connective tissue growth factor expression in scleroderma fibroblasts is dependent on Sp1', *J Biol Chem*, vol. 278, no. 43, pp. 41728-33.

- Holmes, A., Abraham, D. J., Sa, S., Shiwen, X., Black, C. M. & Leask, A. 2001, 'CTGF and SMADs, maintenance of scleroderma phenotype is independent of SMAD signaling', *J Biol Chem*, vol. 276, no. 14, pp. 10594-601.
- Hong, D., Gupta, R., Ancliff, P., Atzberger, A., Brown, J., Soneji, S., Green, J., Colman, S., Piacibello, W., Buckle, V., Tsuzuki, S., Greaves, M. & Enver, T. 2008, 'Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia', *Science*, vol. 319, no. 5861, pp. 336-9.
- Hong, K. H., Yoo, S. A., Kang, S. S., Choi, J. J., Kim, W. U. & Cho, C. S. 2006, 'Hypoxia induces expression of connective tissue growth factor in scleroderma skin fibroblasts', *Clin Exp Immunol*, vol. 146, no. 2, pp. 362-70.
- Honorio, S., Agathangelou, A., Wernert, N., Rothe, M., Maher, E. R. & Latif, F. 2003, 'Frequent epigenetic inactivation of the RASSF1A tumour suppressor gene in testicular tumours and distinct methylation profiles of seminoma and nonseminoma testicular germ cell tumours', *Oncogene*, vol. 22, no. 3, pp. 461-6.
- Hoshijima, M., Hattori, T., Inoue, M., Araki, D., Hanagata, H., Miyauchi, A. & Takigawa, M. 2006, 'CT domain of CCN2/CTGF directly interacts with fibronectin and enhances cell adhesion of chondrocytes through integrin alpha5beta1', *FEBS Lett*, vol. 580, no. 5, pp. 1376-82.
- Hosokawa, Y., Maeda, Y. & Seto, M. 1999, 'Human Helios, an Ikaros-related zinc finger DNA binding protein: cDNA cloning and tissue expression pattern', *Immunogenetics*, vol. 50, no. 1-2, pp. 106-8.
- Houseley, J. & Tollervey, D. 2009, 'The many pathways of RNA degradation', *Cell*, vol. 136, no. 4, pp. 763-76.
- Huang, B. L., Dornbach, L. M. & Lyons, K. M. 2007, 'The 5' untranslated regions (UTRs) of CCN1, CCN2, and CCN4 exhibit cryptic promoter activity', *J Cell Commun Signal*, vol. 1, no. 1, pp. 17-32.
- Huang, E. J. & Reichardt, L. F. 2003, 'Trk receptors: roles in neuronal signal transduction', *Annu Rev Biochem*, vol. 72, pp. 609-42.
- Iacobucci, I., Lonetti, A., Messa, F., Cilloni, D., Arruga, F., Ottaviani, E., Paolini, S., Papayannidis, C., Piccaluga, P. P., Giannoulia, P., Soverini, S., Amabile, M., Poerio, A., Saglio, G., Pane, F., Berton, G., Baruzzi, A., Vitale, A., Chiaretti, S., Perini, G., Foa, R., Baccarani, M. & Martinelli, G. 2008, 'Expression of spliced oncogenic Ikaros isoforms in Philadelphia-positive acute lymphoblastic leukemia patients treated with tyrosine kinase inhibitors: implications for a new mechanism of resistance', *Blood*, vol. 112, no. 9, pp. 3847-55.
- Igarashi, A., Nashiro, K., Kikuchi, K., Sato, S., Ihn, H., Grotendorst, G. R. & Takehara, K. 1995, 'Significant correlation between connective tissue growth factor

gene expression and skin sclerosis in tissue sections from patients with systemic sclerosis', *J Invest Dermatol*, vol. 105, no. 2, pp. 280-4.

Igarashi, A., Okochi, H., Bradham, D. M. & Grotendorst, G. R. 1993, 'Regulation of connective tissue growth factor gene expression in human skin fibroblasts and during wound repair', *Mol Biol Cell*, vol. 4, no. 6, pp. 637-45.

Inoki, I., Shiomi, T., Hashimoto, G., Enomoto, H., Nakamura, H., Makino, K., Ikeda, E., Takata, S., Kobayashi, K. & Okada, Y. 2002, 'Connective tissue growth factor binds vascular endothelial growth factor (VEGF) and inhibits VEGF-induced angiogenesis', *FASEB J*, vol. 16, no. 2, pp. 219-21.

Ishida, Y., Sakamoto, N., Kamibeppu, K., Kakee, N., Iwai, T., Ozono, S., Maeda, N., Okamura, J., Asami, K., Inada, H., Honda, M. & Horibe, K. 2010, 'Late effects and quality of life of childhood cancer survivors: Part 2. Impact of radiotherapy', *Int J Hematol*, vol. 92, no. 1, pp. 95-104.

Ito, T. K., Ishii, G., Chiba, H. & Ochiai, A. 2007, 'The VEGF angiogenic switch of fibroblasts is regulated by MMP-7 from cancer cells', *Oncogene*, vol. 26, no. 51, pp. 7194-203.

Ivkovic, S., Yoon, B. S., Popoff, S. N., Safadi, F. F., Libuda, D. E., Stephenson, R. C., Daluiski, A. & Lyons, K. M. 2003, 'Connective tissue growth factor coordinates chondrogenesis and angiogenesis during skeletal development', *Development*, vol. 130, no. 12, pp. 2779-91.

Jedsadayanmata, A., Chen, C. C., Kireeva, M. L., Lau, L. F. & Lam, S. C. 1999, 'Activation-dependent adhesion of human platelets to Cyr61 and Fisp12/mouse connective tissue growth factor is mediated through integrin alpha(IIb)beta(3)', *J Biol Chem*, vol. 274, no. 34, pp. 24321-7.

Jeha, S., Pei, D., Raimondi, S. C., Onciu, M., Campana, D., Cheng, C., Sandlund, J. T., Ribeiro, R. C., Rubnitz, J. E., Howard, S. C., Downing, J. R., Evans, W. E., Relling, M. V. & Pui, C. H. 2009, 'Increased risk for CNS relapse in pre-B cell leukemia with the t(1;19)/TCF3-PBX1', *Leukemia*, vol. 23, no. 8, pp. 1406-9.

Jude, C. D., Climer, L., Xu, D., Artinger, E., Fisher, J. K. & Ernst, P. 2007, 'Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors', *Cell Stem Cell*, vol. 1, no. 3, pp. 324-37.

Kaatsch, P. 2010, 'Epidemiology of childhood cancer', *Cancer Treat Rev*, vol. 36, no. 4, pp. 277-85.

Kalluri, R. & Zeisberg, M. 2006, 'Fibroblasts in cancer', *Nat Rev Cancer*, vol. 6, no. 5, pp. 392-401.

Kang, H., Chen, I. M., Wilson, C. S., Bedrick, E. J., Harvey, R. C., Atlas, S. R., Devidas, M., Mullighan, C. G., Wang, X., Murphy, M., Ar, K., Wharton, W.,

Borowitz, M. J., Bowman, W. P., Bhojwani, D., Carroll, W. L., Camitta, B. M., Reaman, G. H., Smith, M. A., Downing, J. R., Hunger, S. P. & Willman, C. L. 2010, 'Gene expression classifiers for relapse-free survival and minimal residual disease improve risk classification and outcome prediction in pediatric B-precursor acute lymphoblastic leukemia', *Blood*, vol. 115, no. 7, pp. 1394-405.

Kang, Y., Siegel, P. M., Shu, W., Drobnjak, M., Kakonen, S. M., Cordon-Cardo, C., Guise, T. A. & Massague, J. 2003, 'A multigenic program mediating breast cancer metastasis to bone', *Cancer Cell*, vol. 3, no. 6, pp. 537-49.

Kawaguchi, Y., Ota, Y., Kawamoto, M., Ito, I., Tsuchiya, N., Sugiura, T., Katsumata, Y., Soejima, M., Sato, S., Hasegawa, M., Fujimoto, M., Takehara, K., Kuwana, M., Yamanaka, H. & Hara, M. 2009, 'Association study of a polymorphism of the CTGF gene and susceptibility to systemic sclerosis in the Japanese population', *Ann Rheum Dis*, vol. 68, no. 12, pp. 1921-4.

Kees, U. R. 1987, 'Resistance to 1-beta-D-arabinofuranosylcytosine after high-dose treatment childhood lymphoblastic leukemia: isolation of a drug resistant and a sensitive cell line', *Cancer Res*, vol. 47, no. 12, pp. 3088-91.

Kees, U. R., Campbell, L. J., Ford, J., Willoughby, M. L., Peroni, S. E., Ranford, P. R. & Garson, O. M. 1995, 'New translocation t(2;13)(p12;q34) and rearrangement of the MLL gene in a childhood leukemia cell line', *Genes Chromosomes Cancer*, vol. 12, no. 3, pp. 201-8.

Kees, U. R., Ford, J., Price, P. J., Meyer, B. F. & Herrmann, R. P. 1987, 'PER-117: a new human ALL cell line with an immature thymic phenotype', *Leuk Res*, vol. 11, no. 5, pp. 489-98.

Kees, U. R., Ford, J., Watson, M., Murch, A., Ringner, M., Walker, R. L. & Meltzer, P. 2003, 'Gene expression profiles in a panel of childhood leukemia cell lines mirror critical features of the disease', *Mol Cancer Ther*, vol. 2, no. 7, pp. 671-7.

Kees, U. R., Lukeis, R., Ford, J., Willoughby, M. L. & Garson, O. M. 1990, 'Establishment and characterization of a childhood pre-B acute lymphoblastic leukemia cell line, PER-278, with chromosome translocations t(1;19) and t(1;9)', *Cancer Genet Cytogenet*, vol. 46, no. 2, pp. 201-8.

Kikuchi, R., Tsuda, H., Kanai, Y., Kasamatsu, T., Sengoku, K., Hirohashi, S., Inazawa, J. & Imoto, I. 2007, 'Promoter hypermethylation contributes to frequent inactivation of a putative conditional tumor suppressor gene connective tissue growth factor in ovarian cancer', *Cancer Res*, vol. 67, no. 15, pp. 7095-105.

Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. 1997, 'Identification of a family of low-affinity insulin-like growth factor binding proteins (IGFBPs): characterization of connective tissue growth factor as a

member of the IGFBP superfamily', *Proc Natl Acad Sci U S A*, vol. 94, no. 24, pp. 12981-6.

Kim, J., Sif, S., Jones, B., Jackson, A., Koipally, J., Heller, E., Winandy, S., Viel, A., Sawyer, A., Ikeda, T., Kingston, R. & Georgopoulos, K. 1999, 'Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes', *Immunity*, vol. 10, no. 3, pp. 345-55.

Kim, M., Yim, S. H., Cho, N. S., Kang, S. H., Ko, D. H., Oh, B., Kim, T. Y., Min, H. J., She, C. J., Kang, H. J., Shin, H. Y., Ahn, H. S., Yoon, S. S., Kim, B. K., Shin, H. R., Han, K. S., Cho, H. I. & Lee, D. S. 2009, 'Homozygous deletion of CDKN2A (p16, p14) and CDKN2B (p15) genes is a poor prognostic factor in adult but not in childhood B-lineage acute lymphoblastic leukemia: a comparative deletion and hypermethylation study', *Cancer Genet Cytogenet*, vol. 195, no. 1, pp. 59-65.

Koipally, J., Kim, J., Jones, B., Jackson, A., Avitahl, N., Winandy, S., Trevisan, M., Nichogiannopoulou, A., Kelley, C. & Georgopoulos, K. 1999, 'Ikaros chromatin remodeling complexes in the control of differentiation of the hemo-lymphoid system', *Cold Spring Harb Symp Quant Biol*, vol. 64, pp. 79-86.

Koliopanos, A., Friess, H., di Mola, F. F., Tang, W. H., Kubulus, D., Brigstock, D., Zimmermann, A. & Buchler, M. W. 2002, 'Connective tissue growth factor gene expression alters tumor progression in esophageal cancer', *World J Surg*, vol. 26, no. 4, pp. 420-7.

Komorowsky, C., Ocker, M. & Goppelt-Struebe, M. 2009, 'Differential regulation of connective tissue growth factor in renal cells by histone deacetylase inhibitors', *J Cell Mol Med*, vol. 13, no. 8B, pp. 2353-64.

Kondo, S., Kubota, S., Mukudai, Y., Moritani, N., Nishida, T., Matsushita, H., Matsumoto, S., Sugahara, T. & Takigawa, M. 2006, 'Hypoxic regulation of stability of connective tissue growth factor/CCN2 mRNA by 3'-untranslated region interacting with a cellular protein in human chondrosarcoma cells', *Oncogene*, vol. 25, no. 7, pp. 1099-110.

Kondo, Y. & Issa, J. P. 2010, 'DNA methylation profiling in cancer', *Expert Rev Mol Med*, vol. 12, p. e23.

Kouzarides, T. 2007, 'Chromatin modifications and their function', *Cell*, vol. 128, no. 4, pp. 693-705.

Kriaucionis, S. & Heintz, N. 2009, 'The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain', *Science*, vol. 324, no. 5929, pp. 929-30.

Kroening, S., Neubauer, E., Wessel, J., Wiesener, M. & Goppelt-Struebe, M. 2009, 'Hypoxia interferes with connective tissue growth factor (CTGF) gene expression in

human proximal tubular cell lines', *Nephrol Dial Transplant*, vol. 24, no. 11, pp. 3319-25.

Krug, U., Ganser, A. & Koeffler, H. P. 2002, 'Tumor suppressor genes in normal and malignant hematopoiesis', *Oncogene*, vol. 21, no. 21, pp. 3475-95.

Kubo, M., Kikuchi, K., Nashiro, K., Kakinuma, T., Hayashi, N., Nanko, H. & Tamaki, K. 1998, 'Expression of fibrogenic cytokines in desmoplastic malignant melanoma', *Br J Dermatol*, vol. 139, no. 2, pp. 192-7.

Kubota, S., Kondo, S., Eguchi, T., Hattori, T., Nakanishi, T., Pomerantz, R. J. & Takigawa, M. 2000, 'Identification of an RNA element that confers post-transcriptional repression of connective tissue growth factor/hypertrophic chondrocyte specific 24 (ctgf/hcs24) gene: similarities to retroviral RNA-protein interactions', *Oncogene*, vol. 19, no. 41, pp. 4773-86.

Kubota, S., Mukudai, Y., Moritani, N. H., Nakao, K., Kawata, K. & Takigawa, M. 2005, 'Translational repression by the cis-acting element of structure-anchored repression (CAESAR) of human ctgf/ccn2 mRNA', *FEBS Lett*, vol. 579, no. 17, pp. 3751-8.

Kuiper, E. J., Hughes, J. M., Van Geest, R. J., Vogels, I. M., Goldschmeding, R., Van Noorden, C. J., Schlingemann, R. O. & Klaassen, I. 2007a, 'Effect of VEGF-A on expression of profibrotic growth factor and extracellular matrix genes in the retina', *Invest Ophthalmol Vis Sci*, vol. 48, no. 9, pp. 4267-76.

Kuiper, R. P., Schoenmakers, E. F., van Reijmersdal, S. V., Hehir-Kwa, J. Y., van Kessel, A. G., van Leeuwen, F. N. & Hoogerbrugge, P. M. 2007b, 'High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression', *Leukemia*, vol. 21, no. 6, pp. 1258-66.

Kuiper, R. P., Waanders, E., van der Velden, V. H., van Reijmersdal, S. V., Venkatachalam, R., Scheijen, B., Sonneveld, E., van Dongen, J. J., Veerman, A. J., van Leeuwen, F. N., van Kessel, A. G. & Hoogerbrugge, P. M. 2010, 'IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL', *Leukemia*, vol. 24, no. 7, pp. 1258-64.

Kundel, D. W., Brecher, G., Bodey, G. P. & Brittin, G. M. 1964, 'Reticulin Fibrosis and Bone Infarction in Acute Leukemia. Implications for Prognosis', *Blood*, vol. 23, pp. 526-44.

Lane, S. W., Scadden, D. T. & Gilliland, D. G. 2009, 'The leukemic stem cell niche: current concepts and therapeutic opportunities', *Blood*, vol. 114, no. 6, pp. 1150-7.

Lang, C., Sauter, M., Szalay, G., Racchi, G., Grassi, G., Rainaldi, G., Mercatanti, A., Lang, F., Kandolf, R. & Klingel, K. 2007, 'Connective tissue growth factor: a

crucial cytokine-mediating cardiac fibrosis in ongoing enterovirus myocarditis', *J Mol Med*.

Laurila, K. & Lahdesmaki, H. 2009, 'Systematic analysis of disease-related regulatory mutation classes reveals distinct effects on transcription factor binding', *In Silico Biol*, vol. 9, no. 4, pp. 209-24.

Leask, A. 2004, 'Transcriptional profiling of the scleroderma fibroblast reveals a potential role for connective tissue growth factor (CTGF) in pathological fibrosis', *Keio J Med*, vol. 53, no. 2, pp. 74-7.

Leask, A. 2010, 'Potential therapeutic targets for cardiac fibrosis: TGFbeta, angiotensin, endothelin, CCN2, and PDGF, partners in fibroblast activation', *Circ Res*, vol. 106, no. 11, pp. 1675-80.

Leask, A. & Abraham, D. J. 2006, 'All in the CCN family: essential matricellular signaling modulators emerge from the bunker', *J Cell Sci*, vol. 119, no. Pt 23, pp. 4803-10.

Lee, C. H., Shah, B., Muioli, E. K. & Mao, J. J. 2010, 'CTGF directs fibroblast differentiation from human mesenchymal stem/stromal cells and defines connective tissue healing in a rodent injury model', *J Clin Invest*, vol. 120, no. 9, pp. 3340-9.

Lee, J. H. & Skalnik, D. G. 2005, 'CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex', *J Biol Chem*, vol. 280, no. 50, pp. 41725-31.

Li, B., Jackson, J., Simon, M. D., Fleharty, B., Gogol, M., Seidel, C., Workman, J. L. & Shilatifard, A. 2009a, 'Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription', *J Biol Chem*, vol. 284, no. 12, pp. 7970-6.

Li, X., Zhao, L., Jiang, H. & Wang, W. 2009b, 'Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes', *J Mol Evol*, vol. 68, no. 1, pp. 56-65.

Li, Y., Lu, Y., Ceng, Y. & Yang, X. 2007, '[Expression of connective tissue growth factor (CTGF), osteopontin (OPN) and clinical significances in the laryngeal squamous cell carcinoma tissues]', *Lin Chung Er Bi Yan Hou Tou Jing Wai Ke Za Zhi*, vol. 21, no. 3, pp. 121-3.

Lin, B. R., Chang, C. C., Che, T. F., Chen, S. T., Chen, R. J., Yang, C. Y., Jeng, Y. M., Liang, J. T., Lee, P. H., Chang, K. J., Chau, Y. P. & Kuo, M. L. 2005, 'Connective tissue growth factor inhibits metastasis and acts as an independent prognostic marker in colorectal cancer', *Gastroenterology*, vol. 128, no. 1, pp. 9-23.

Lind, G. E., Thorstensen, L., Lovig, T., Meling, G. I., Hamelin, R., Rognum, T. O., Esteller, M. & Lothe, R. A. 2004, 'A CpG island hypermethylation profile of primary colorectal carcinomas and colon cancer cell lines', *Mol Cancer*, vol. 3, p. 28.

Lionetti, M., Biasiolo, M., Agnelli, L., Todoerti, K., Mosca, L., Fabris, S., Sales, G., Deliliers, G. L., Bicciato, S., Lombardi, L., Bortoluzzi, S. & Neri, A. 2009, 'Identification of microRNA expression patterns and definition of a microRNA/mRNA regulatory network in distinct molecular groups of multiple myeloma', *Blood*, vol. 114, no. 25, pp. e20-6.

Lo Celso, C., Fleming, H. E., Wu, J. W., Zhao, C. X., Miake-Lye, S., Fujisaki, J., Cote, D., Rowe, D. W., Lin, C. P. & Scadden, D. T. 2009, 'Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche', *Nature*, vol. 457, no. 7225, pp. 92-6.

Luscher, B. & Larsson, L. G. 1999, 'The basic region/helix-loop-helix/leucine zipper domain of Myc proto-oncoproteins: function and regulation', *Oncogene*, vol. 18, no. 19, pp. 2955-66.

Maia, A. T., van der Velden, V. H., Harrison, C. J., Szczepanski, T., Williams, M. D., Griffiths, M. J., van Dongen, J. J. & Greaves, M. F. 2003, 'Prenatal origin of hyperdiploid acute lymphoblastic leukemia in identical twins', *Leukemia*, vol. 17, no. 11, pp. 2202-6.

Manetti, M., Neumann, E., Milia, A. F., Tarner, I. H., Bechi, P., Matucci-Cerinic, M., Ibba-Manneschi, L. & Muller-Ladner, U. 2007, 'Severe fibrosis and increased expression of fibrogenic cytokines in the gastric wall of systemic sclerosis patients', *Arthritis Rheum*, vol. 56, no. 10, pp. 3442-3447.

Mansson, R., Tsapogas, P., Akerlund, M., Lagergren, A., Gisler, R. & Sigvardsson, M. 2004, 'Pearson correlation analysis of microarray data allows for the identification of genetic targets for early B-cell factor', *J Biol Chem*, vol. 279, no. 17, pp. 17905-13.

Mansson, R., Zandi, S., Bryder, D. & Sigvardsson, M. 2009, 'The Road to Commitment: Lineage Restriction Events in Hematopoiesis', in *Molecular Basis of Hematopoiesis*, eds A. Wickrema & B. Kee, Springer, New York, pp. 23-46.

Markl, I. D., Cheng, J., Liang, G., Shibata, D., Laird, P. W. & Jones, P. A. 2001, 'Global and gene-specific epigenetic patterns in human bladder cancer genomes are relatively stable in vivo and in vitro over time', *Cancer Res*, vol. 61, no. 15, pp. 5875-84.

Marks, D. I., Kurz, B. W., Link, M. P., Ng, E., Shuster, J. J., Lauer, S. J., Carroll, D., Brodsky, I. & Haines, D. S. 1997, 'Altered expression of p53 and mdm-2 proteins at diagnosis is associated with early treatment failure in childhood acute lymphoblastic leukemia', *J Clin Oncol*, vol. 15, no. 3, pp. 1158-62.

Martin-Subero, J. I., Ammerpohl, O., Bibikova, M., Wickham-Garcia, E., Agirre, X., Alvarez, S., Bruggemann, M., Bug, S., Calasanz, M. J., Deckert, M., Dreyling, M., Du, M. Q., Durig, J., Dyer, M. J., Fan, J. B., Gesk, S., Hansmann, M. L., Harder, L., Hartmann, S., Klapper, W., Kuppers, R., Montesinos-Rongen, M., Nagel, I., Pott, C., Richter, J., Roman-Gomez, J., Seifert, M., Stein, H., Suela, J., Trumper, L., Vater, I., Prosper, F., Haferlach, C., Cruz Cigudosa, J. & Siebert, R. 2009, 'A comprehensive microarray-based DNA methylation study of 367 hematological neoplasms', *PLoS One*, vol. 4, no. 9, p. e6986.

Martinerie, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. 1992, 'Physical mapping of human loci homologous to the chicken *nov* proto-oncogene', *Oncogene*, vol. 7, no. 12, pp. 2529-34.

Mayr, C. & Bartel, D. P. 2009, 'Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells', *Cell*, vol. 138, no. 4, pp. 673-84.

McNally, R. J. & Parker, L. 2006, 'Environmental factors and childhood acute leukemias and lymphomas', *Leuk Lymphoma*, vol. 47, no. 4, pp. 583-98.

McNeer, J. L. & Nachman, J. B. 2010, 'The optimal use of steroids in paediatric acute lymphoblastic leukaemia: no easy answers', *Br J Haematol*, vol. 149, no. 5, pp. 638-52.

McNeil, D. E., Cote, T. R., Clegg, L. & Mauer, A. 2002, 'SEER update of incidence and trends in pediatric malignancies: acute lymphoblastic leukemia', *Med Pediatr Oncol*, vol. 39, no. 6, pp. 554-7; discussion 552-3.

Meads, M. B., Hazlehurst, L. A. & Dalton, W. S. 2008, 'The bone marrow microenvironment as a tumor sanctuary and contributor to drug resistance', *Clin Cancer Res*, vol. 14, no. 9, pp. 2519-26.

Meleshko, A. N., Movchan, L. V., Belevtsev, M. V. & Savitskaja, T. V. 2008, 'Relative expression of different Ikaros isoforms in childhood acute leukemia', *Blood Cells Mol Dis*, vol. 41, no. 3, pp. 278-83.

Melrose, J., Hayes, A. J., Whitelock, J. M. & Little, C. B. 2008, 'Perlecan, the "jack of all trades" proteoglycan of cartilaginous weight-bearing connective tissues', *Bioessays*, vol. 30, no. 5, pp. 457-69.

Mendez-Ferrer, S., Michurina, T. V., Ferraro, F., Mazloom, A. R., Macarthur, B. D., Lira, S. A., Scadden, D. T., Ma'ayan, A., Enikolopov, G. N. & Frenette, P. S. 2010, 'Mesenchymal and haematopoietic stem cells form a unique bone marrow niche', *Nature*, vol. 466, no. 7308, pp. 829-34.

Mercurio, S., Latinkic, B., Itasaki, N., Krumlauf, R. & Smith, J. C. 2004, 'Connective-tissue growth factor modulates WNT signalling and interacts with the WNT receptor complex', *Development*, vol. 131, no. 9, pp. 2137-47.

Messeguer, X., Escudero, R., Farre, D., Nunez, O., Martinez, J. & Alba, M. M. 2002, 'PROMO: detection of known transcription regulatory elements using species-tailored searches', *Bioinformatics*, vol. 18, no. 2, pp. 333-4.

Meyer, C., Kowarz, E., Hofmann, J., Renneville, A., Zuna, J., Trka, J., Ben Abdelali, R., Macintyre, E., De Braekeleer, E., De Braekeleer, M., Delabesse, E., de Oliveira, M. P., Cave, H., Clappier, E., van Dongen, J. J., Balgobind, B. V., van den Heuvel-Eibrink, M. M., Beverloo, H. B., Panzer-Grumayer, R., Teigler-Schlegel, A., Harbott, J., Kjeldsen, E., Schnittger, S., Koehl, U., Gruhn, B., Heidenreich, O., Chan, L. C., Yip, S. F., Krzywinski, M., Eckert, C., Moricke, A., Schrappe, M., Alonso, C. N., Schafer, B. W., Krauter, J., Lee, D. A., Zur Stadt, U., Te Kronnie, G., Sutton, R., Izraeli, S., Trakhtenbrot, L., Lo Nigro, L., Tsaour, G., Fechina, L., Szczepanski, T., Strehl, S., Ilencikova, D., Molkentin, M., Burmeister, T., Dinger, T., Klingebiel, T. & Marschalek, R. 2009, 'New insights to the MLL recombinome of acute leukemias', *Leukemia*, vol. 23, no. 8, pp. 1490-9.

Milani, L., Lundmark, A., Kiialainen, A., Nordlund, J., Flaegstad, T., Forestier, E., Heyman, M., Jonmundsson, G., Kanerva, J., Schmiegelow, K., Soderhall, S., Gustafsson, M. G., Lonnerholm, G. & Syvanen, A. C. 2010, 'DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia', *Blood*, vol. 115, no. 6, pp. 1214-25.

Mitelman, F., Johansson, B. & Mertens, F. 2010, *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (2010)*. Accessed: Jan 2010

Molnar, A. & Georgopoulos, K. 1994, 'The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins', *Mol Cell Biol*, vol. 14, no. 12, pp. 8292-303.

Moritani, N. H., Kubota, S., Nishida, T., Kawaki, H., Kondo, S., Sugahara, T. & Takigawa, M. 2003, 'Suppressive effect of overexpressed connective tissue growth factor on tumor cell growth in a human oral squamous cell carcinoma-derived cell line', *Cancer Lett*, vol. 192, no. 2, pp. 205-14.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. 2008, 'Mapping and quantifying mammalian transcriptomes by RNA-Seq', *Nat Methods*, vol. 5, no. 7, pp. 621-8.

Muehlich, S., Cicha, I., Garlich, C. D., Krueger, B., Posern, G. & Goppelt-Strube, M. 2007, 'Actin-dependent regulation of connective tissue growth factor', *Am J Physiol Cell Physiol*, vol. 292, no. 5, pp. C1732-8.

Mullighan, C. G. & Downing, J. R. 2009, 'Genome-wide profiling of genetic alterations in acute lymphoblastic leukemia: recent insights and future directions', *Leukemia*, vol. 23, no. 7, pp. 1209-18.

Mullighan, C. G., Goorha, S., Radtke, I., Miller, C. B., Coustan-Smith, E., Dalton, J. D., Girtman, K., Mathew, S., Ma, J., Pounds, S. B., Su, X., Pui, C. H., Relling, M.

V., Evans, W. E., Shurtleff, S. A. & Downing, J. R. 2007, 'Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia', *Nature*, vol. 446, no. 7137, pp. 758-64.

Mullighan, C. G., Miller, C. B., Radtke, I., Phillips, L. A., Dalton, J., Ma, J., White, D., Hughes, T. P., Le Beau, M. M., Pui, C. H., Relling, M. V., Shurtleff, S. A. & Downing, J. R. 2008, 'BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros', *Nature*, vol. 453, no. 7191, pp. 110-4.

Mullighan, C. G., Su, X., Zhang, J., Radtke, I., Phillips, L. A., Miller, C. B., Ma, J., Liu, W., Cheng, C., Schulman, B. A., Harvey, R. C., Chen, I. M., Clifford, R. J., Carroll, W. L., Reaman, G., Bowman, W. P., Devidas, M., Gerhard, D. S., Yang, W., Relling, M. V., Shurtleff, S. A., Campana, D., Borowitz, M. J., Pui, C. H., Smith, M., Hunger, S. P., Willman, C. L. & Downing, J. R. 2009, 'Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia', *N Engl J Med*, vol. 360, no. 5, pp. 470-80.

Murphy, G. & Nagase, H. 2008, 'Progress in matrix metalloproteinase research', *Mol Aspects Med*, vol. 29, no. 5, pp. 290-308.

Nagasawa, T. 2006, 'Microenvironmental niches in the bone marrow required for B-cell development', *Nat Rev Immunol*, vol. 6, no. 2, pp. 107-16.

Nagatoshi, Y., Matsuzaki, A., Suminoe, A., Inada, H., Ueda, K., Kawakami, K., Yanai, F., Nakayama, H., Moritake, H., Itonaga, N., Hotta, N., Fujita, K., Hidaka, Y., Yamanaka, T., Kawano, Y. & Okamura, J. 2010, 'Randomized trial to compare LSA2L2-type maintenance therapy to daily 6-mercaptopurine and weekly methotrexate with vincristine and dexamethasone pulse for children with acute lymphoblastic leukemia', *Pediatr Blood Cancer*, vol. 55, no. 2, pp. 239-47.

Ness, K. K., Morris, E. B., Nolan, V. G., Howell, C. R., Gilchrist, L. S., Stovall, M., Cox, C. L., Klosky, J. L., Gajjar, A. & Neglia, J. P. 2010, 'Physical performance limitations among adult survivors of childhood brain tumors', *Cancer*, vol. 116, no. 12, pp. 3034-44.

Nestor, C., Ruzov, A., Meehan, R. & Dunican, D. 2010, 'Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA', *Biotechniques*, vol. 48, no. 4, pp. 317-9.

Nguyen, T. Q., Roestenberg, P., van Nieuwenhoven, F. A., Bovenschen, N., Li, Z., Xu, L., Oliver, N., Aten, J., Joles, J. A., Vial, C., Brandan, E., Lyons, K. M. & Goldschmeding, R. 2008, 'CTGF inhibits BMP-7 signaling in diabetic nephropathy', *J Am Soc Nephrol*, vol. 19, no. 11, pp. 2098-107.

Nie, L., Wu, G. & Zhang, W. 2006, 'Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in *Desulfovibrio vulgaris*: a quantitative analysis', *Genetics*, vol. 174, no. 4, pp. 2229-43.

- Nielsen, J. S. & McNagny, K. M. 2009, 'CD34 is a key regulator of hematopoietic stem cell trafficking to bone marrow and mast cell progenitor trafficking in the periphery', *Microcirculation*, vol. 16, no. 6, pp. 487-96.
- Niitsu, N., Hayashi, Y., Sugita, K. & Honma, Y. 2001, 'Sensitization by 5-aza-2'-deoxycytidine of leukaemia cells with MLL abnormalities to induction of differentiation by all-trans retinoic acid and 1alpha,25-dihydroxyvitamin D3', *Br J Haematol*, vol. 112, no. 2, pp. 315-26.
- Nishida, T., Kawaki, H., Baxter, R. M., Deyoung, R. A., Takigawa, M. & Lyons, K. M. 2007, 'CCN2 (Connective Tissue Growth Factor) is essential for extracellular matrix production and integrin signaling in chondrocytes', *J Cell Commun Signal*, vol. 1, no. 1, pp. 45-58.
- Nishida, T., Maeda, A., Kubota, S. & Takigawa, M. 2008, 'Role of mechanical-stress inducible protein Hcs24/CTGF/CCN2 in cartilage growth and regeneration: mechanical stress induces expression of Hcs24/CTGF/CCN2 in a human chondrocytic cell line HCS-2/8, rabbit costal chondrocytes and meniscus tissue cells', *Biorheology*, vol. 45, no. 3-4, pp. 289-99.
- Nishida, T., Nakanishi, T., Asano, M., Shimo, T. & Takigawa, M. 2000, 'Effects of CTGF/Hcs24, a hypertrophic chondrocyte-specific gene product, on the proliferation and differentiation of osteoblastic cells in vitro', *J Cell Physiol*, vol. 184, no. 2, pp. 197-206.
- Noren-Nystrom, U., Roos, G., Bergh, A., Botling, J., Lonnerholm, G., Porwit, A., Heyman, M. & Forestier, E. 2008, 'Bone marrow fibrosis in childhood acute lymphoblastic leukemia correlates to biological factors, treatment response and outcome', *Leukemia*, vol. 22, no. 3, pp. 504-10.
- Oeffinger, K. C., Nathan, P. C. & Kremer, L. C. 2008, 'Challenges after curative treatment for childhood cancer and long-term follow up of survivors', *Pediatr Clin North Am*, vol. 55, no. 1, pp. 251-73, xiii.
- Ohgawara, T., Kubota, S., Kawaki, H., Kondo, S., Eguchi, T., Kurio, N., Aoyama, E., Sasaki, A. & Takigawa, M. 2009, 'Regulation of chondrocytic phenotype by micro RNA 18a: involvement of Ccn2/Ctgf as a major target gene', *FEBS Lett*, vol. 583, no. 6, pp. 1006-10.
- Okitsu, C. Y. & Hsieh, C. L. 2007, 'DNA methylation dictates histone H3K4 methylation', *Mol Cell Biol*, vol. 27, no. 7, pp. 2746-57.
- Okumura, A. J., Peterson, L. F., Lo, M. C. & Zhang, D. E. 2007, 'Expression of AML/Runx and ETO/MTG family members during hematopoietic differentiation of embryonic stem cells', *Exp Hematol*, vol. 35, no. 6, pp. 978-88.

Omatsu, Y., Sugiyama, T., Kohara, H., Kondoh, G., Fujii, N., Kohno, K. & Nagasawa, T. 2010, 'The essential functions of adipo-osteogenic progenitors as the hematopoietic stem and progenitor cell niche', *Immunity*, vol. 33, no. 3, pp. 387-99.

Onciu, M. 2009, 'Acute lymphoblastic leukemia', *Hematol Oncol Clin North Am*, vol. 23, no. 4, pp. 655-74.

Park, J., Kunjibettu, S., McMahon, S. B. & Cole, M. D. 2001, 'The ATM-related domain of TRRAP is required for histone acetyltransferase recruitment and Myc-dependent oncogenesis', *Genes Dev*, vol. 15, no. 13, pp. 1619-24.

Parrish, Y. K., Baez, I., Milford, T. A., Benitez, A., Galloway, N., Rogerio, J. W., Sahakian, E., Kagoda, M., Huang, G., Hao, Q. L., Sevilla, Y., Barsky, L. W., Zielinska, E., Price, M. A., Wall, N. R., Dovat, S. & Payne, K. J. 2009, 'IL-7 Dependence in human B lymphopoiesis increases during progression of ontogeny from cord blood to bone marrow', *J Immunol*, vol. 182, no. 7, pp. 4255-66.

Paulsson, K. & Johansson, B. 2009, 'High hyperdiploid childhood acute lymphoblastic leukemia', *Genes Chromosomes Cancer*, vol. 48, no. 8, pp. 637-60.

Peltz, S. W., He, F., Welch, E. & Jacobson, A. 1994, 'Nonsense-mediated mRNA decay in yeast', *Prog Nucleic Acid Res Mol Biol*, vol. 47, pp. 271-98.

Perbal, B. 2001, 'NOV (nephroblastoma overexpressed) and the CCN family of genes: structural and functional issues', *Mol Pathol*, vol. 54, no. 2, pp. 57-79.

Perbal, B. 2009, 'Alternative splicing of CCN mRNAs it has been upon us', *J Cell Commun Signal*, vol. 3, no. 2, pp. 153-7.

Perdomo, J., Holmes, M., Chong, B. & Crossley, M. 2000, 'Eos and pegasus, two members of the Ikaros family of proteins with distinct DNA binding activities', *J Biol Chem*, vol. 275, no. 49, pp. 38347-54.

Phanish, M. K., Wahab, N. A., Colville-Nash, P., Hendry, B. M. & Dockrell, M. E. 2006, 'The differential role of Smad2 and Smad3 in the regulation of pro-fibrotic TGFbeta1 responses in human proximal-tubule epithelial cells', *Biochem J*, vol. 393, no. Pt 2, pp. 601-7.

Pickles, M. & Leask, A. 2007, 'Analysis of CCN2 promoter activity in PANC-1 cells: regulation by ras/MEK/ERK', *J Cell Commun Signal*, vol. 1, no. 2, pp. 85-90.

Pieters, R. & Carroll, W. L. 2010, 'Biology and treatment of acute lymphoblastic leukemia', *Hematol Oncol Clin North Am*, vol. 24, no. 1, pp. 1-18.

Pomerantz, J., Schreiber-Agus, N., Liegeois, N. J., Silverman, A., Alland, L., Chin, L., Potes, J., Chen, K., Orlov, I., Lee, H. W., Cordon-Cardo, C. & DePinho, R. A. 1998, 'The Ink4a tumor suppressor gene product, p19Arf, interacts with MDM2 and neutralizes MDM2's inhibition of p53', *Cell*, vol. 92, no. 6, pp. 713-23.

Poppe, B., De Paepe, A. & Speleman, F. 2007, 'Acquired chromosomal rearrangements targeting selected transcription factors: contribution of molecular cytogenetic and expression analyses to the identification of clinically and biologically relevant subgroups in hematological malignancies', *Verh K Acad Geneeskd Belg*, vol. 69, no. 1, pp. 47-64.

Portales-Casamar, E., Thongjuea, S., Kwon, A. T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W. W. & Sandelin, A. 2010, 'JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles', *Nucleic Acids Res*, vol. 38, no. Database issue, pp. D105-10.

Pui, C. H. 2006, 'Childhood Leukemias', in, 2nd edn, Cambridge University Press, Cambridge, pp. 166-169.

Pui, C. H., Carroll, W. L., Meshinchi, S. & Arceci, R. J. 2011, 'Biology, risk stratification, and therapy of pediatric acute leukemias: an update', *J Clin Oncol*, vol. 29, no. 5, pp. 551-65.

Pui, C. H. & Evans, W. E. 2006, 'Treatment of acute lymphoblastic leukemia', *N Engl J Med*, vol. 354, no. 2, pp. 166-78.

Pui, C. H., Pei, D., Sandlund, J. T., Ribeiro, R. C., Rubnitz, J. E., Raimondi, S. C., Onciu, M., Campana, D., Kun, L. E., Jeha, S., Cheng, C., Howard, S. C., Metzger, M. L., Bhojwani, D., Downing, J. R., Evans, W. E. & Relling, M. V. 2010, 'Long-term results of St Jude Total Therapy Studies 11, 12, 13A, 13B, and 14 for childhood acute lymphoblastic leukemia', *Leukemia*, vol. 24, no. 2, pp. 371-82.

Pui, C. H., Sandlund, J. T., Pei, D., Rivera, G. K., Howard, S. C., Ribeiro, R. C., Rubnitz, J. E., Razzouk, B. I., Hudson, M. M., Cheng, C., Raimondi, S. C., Behm, F. G., Downing, J. R., Relling, M. V. & Evans, W. E. 2003, 'Results of therapy for acute lymphoblastic leukemia in black and white children', *JAMA*, vol. 290, no. 15, pp. 2001-7.

Purton, L. E. & Scadden, D. T. 2007, 'Limiting factors in murine hematopoietic stem cell assays', *Cell Stem Cell*, vol. 1, no. 3, pp. 263-70.

Raimondi, S. C., Behm, F. G., Roberson, P. K., Williams, D. L., Pui, C. H., Crist, W. M., Look, A. T. & Rivera, G. K. 1990, 'Cytogenetics of pre-B-cell acute lymphoblastic leukemia with emphasis on prognostic implications of the t(1;19)', *J Clin Oncol*, vol. 8, no. 8, pp. 1380-8.

Rebollo, A. & Schmitt, C. 2003, 'Ikaros, Aiolos and Helios: transcription regulators and lymphoid malignancies', *Immunol Cell Biol*, vol. 81, no. 3, pp. 171-5.

Recchia, A. G., Filice, E., Pellegrino, D., Dobrina, A., Cerra, M. C. & Maggiolini, M. 2009, 'Endothelin-1 induces connective tissue growth factor expression in cardiomyocytes', *J Mol Cell Cardiol*, vol. 46, no. 3, pp. 352-9.

Riggs, A. D. 1975, 'X inactivation, differentiation, and DNA methylation', *Cytogenet Cell Genet*, vol. 14, no. 1, pp. 9-25.

Rodriguez-Pinto, D. 2005, 'B cells as antigen presenting cells', *Cell Immunol*, vol. 238, no. 2, pp. 67-75.

Ross, M. E., Zhou, X., Song, G., Shurtleff, S. A., Girtman, K., Williams, W. K., Liu, H. C., Mahfouz, R., Raimondi, S. C., Lenny, N., Patel, A. & Downing, J. R. 2003, 'Classification of pediatric acute lymphoblastic leukemia by gene expression profiling', *Blood*, vol. 102, no. 8, pp. 2951-9.

Rowe, J. M., Buck, G., Burnett, A. K., Chopra, R., Wiernik, P. H., Richards, S. M., Lazarus, H. M., Franklin, I. M., Litzow, M. R., Ciobanu, N., Prentice, H. G., Durrant, J., Tallman, M. S. & Goldstone, A. H. 2005, 'Induction therapy for adults with acute lymphoblastic leukemia: results of more than 1500 patients from the international ALL trial: MRC UKALL XII/ECOG E2993', *Blood*, vol. 106, no. 12, pp. 3760-7.

Ruan, K., Fang, X. & Ouyang, G. 2009, 'MicroRNAs: novel regulators in the hallmarks of human cancer', *Cancer Lett*, vol. 285, no. 2, pp. 116-26.

Rubnitz, J. E. & Crist, W. M. 1997, 'Molecular genetics of childhood cancer: implications for pathogenesis, diagnosis, and treatment', *Pediatrics*, vol. 100, no. 1, pp. 101-8.

Rueda, B., Simeon, C., Hesselstrand, R., Herrick, A., Worthington, J., Ortego-Centeno, N., Riemekasten, G., Fonollosa, V., Vonk, M. C., van den Hoogen, F. H., Sanchez-Roman, J., Aguirre-Zamorano, M. A., Garcia-Portales, R., Pros, A., Camps, M. T., Gonzalez-Gay, M. A., Gonzalez-Escribano, M. F., Coenen, M. J., Lambert, N., Nelson, J. L., Radstake, T. R. & Martin, J. 2009, 'A large multicentre analysis of CTGF -945 promoter polymorphism does not confirm association with systemic sclerosis susceptibility or phenotype', *Ann Rheum Dis*, vol. 68, no. 10, pp. 1618-20.

Ruiz, A., Williams, O. & Brady, H. J. 2008, 'The Ikaros splice isoform, Ikaros 6, immortalizes murine haematopoietic progenitor cells', *Int J Cancer*, vol. 123, no. 6, pp. 1240-5.

Sala-Torra, O., Gundacker, H. M., Stirewalt, D. L., Ladne, P. A., Pogossova-Agadjanya, E. L., Slovak, M. L., Willman, C. L., Heimfeld, S., Boldt, D. H. & Radich, J. P. 2007, 'Connective tissue growth factor (CTGF) expression and outcome in adult patients with acute lymphoblastic leukemia', *Blood*, vol. 109, no. 7, pp. 3080-3.

Samarin, J., Wessel, J., Cicha, I., Kroening, S., Warnecke, C. & Goppelt-Struebe, M. 2010, 'FoxO proteins mediate hypoxic induction of connective tissue growth factor in endothelial cells', *J Biol Chem*, vol. 285, no. 7, pp. 4328-36.

Sarraf, S. A. & Stancheva, I. 2004, 'Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly', *Mol Cell*, vol. 15, no. 4, pp. 595-605.

Sato, S., Nagaoka, T., Hasegawa, M., Tamatani, T., Nakanishi, T., Takigawa, M. & Takehara, K. 2000, 'Serum levels of connective tissue growth factor are elevated in patients with systemic sclerosis: association with extent of skin sclerosis and severity of pulmonary fibrosis', *J Rheumatol*, vol. 27, no. 1, pp. 149-54.

Schober, J. M., Chen, N., Grzeszkiewicz, T. M., Jovanovic, I., Emeson, E. E., Ugarova, T. P., Ye, R. D., Lau, L. F. & Lam, S. C. 2002, 'Identification of integrin alpha(M)beta(2) as an adhesion receptor on peripheral blood monocytes for Cyr61 (CCN1) and connective tissue growth factor (CCN2): immediate-early gene products expressed in atherosclerotic lesions', *Blood*, vol. 99, no. 12, pp. 4457-65.

Schofield, K. P., Gallagher, J. T. & David, G. 1999, 'Expression of proteoglycan core proteins in human bone marrow stroma', *Biochem J*, vol. 343 Pt 3, pp. 663-8.

Segarini, P. R., Nesbitt, J. E., Li, D., Hays, L. G., Yates, J. R., 3rd & Carmichael, D. F. 2001, 'The low density lipoprotein receptor-related protein/alpha2-macroglobulin receptor is a receptor for connective tissue growth factor', *J Biol Chem*, vol. 276, no. 44, pp. 40659-67.

Seifert, H. H., Schmiemann, V., Mueller, M., Kazimirek, M., Onofre, F., Neuhausen, A., Florl, A. R., Ackermann, R., Boecking, A., Schulz, W. A. & Grote, H. J. 2007, 'In situ detection of global DNA hypomethylation in exfoliative urine cytology of patients with suspected bladder cancer', *Exp Mol Pathol*, vol. 82, no. 3, pp. 292-7.

Shakunaga, T., Ozaki, T., Ohara, N., Asaumi, K., Doi, T., Nishida, K., Kawai, A., Nakanishi, T., Takigawa, M. & Inoue, H. 2000, 'Expression of connective tissue growth factor in cartilaginous tumors', *Cancer*, vol. 89, no. 7, pp. 1466-73.

Shi-Wen, X., Leask, A. & Abraham, D. 2008, 'Regulation and function of connective tissue growth factor/CCN2 in tissue repair, scarring and fibrosis', *Cytokine Growth Factor Rev*, vol. 19, no. 2, pp. 133-44.

Shi-wen, X., Stanton, L. A., Kennedy, L., Pala, D., Chen, Y., Howat, S. L., Renzoni, E. A., Carter, D. E., Bou-Gharios, G., Stratton, R. J., Pearson, J. D., Beier, F., Lyons, K. M., Black, C. M., Abraham, D. J. & Leask, A. 2006, 'CCN2 is necessary for adhesive responses to transforming growth factor-beta1 in embryonic fibroblasts', *J Biol Chem*, vol. 281, no. 16, pp. 10715-26.

Shimo, T., Kubota, S., Yoshioka, N., Ibaragi, S., Isowa, S., Eguchi, T., Sasaki, A. & Takigawa, M. 2006, 'Pathogenic role of connective tissue growth factor (CTGF/CCN2) in osteolytic metastasis of breast cancer', *J Bone Miner Res*, vol. 21, no. 7, pp. 1045-59.

Sidhu, S., Martin, E., Gicquel, C., Melki, J., Clark, S. J., Campbell, P., Magarey, C. J., Schulte, K. M., Roher, H. D., Delbridge, L. & Robinson, B. G. 2005, 'Mutation and methylation analysis of TP53 in adrenal carcinogenesis', *Eur J Surg Oncol*, vol. 31, no. 5, pp. 549-54.

Skorski, T. 2002, 'BCR/ABL regulates response to DNA damage: the role in resistance to genotoxic treatment and in genomic instability', *Oncogene*, vol. 21, no. 56, pp. 8591-604.

Smith, E. & Shilatifard, A. 2011, 'The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes', *Mol Cell*, vol. 40, no. 5, pp. 689-701.

Smith, I. M., Glazer, C. A., Mithani, S. K., Ochs, M. F., Sun, W., Bhan, S., Vostrov, A., Abdullaev, Z., Lobanenkov, V., Gray, A., Liu, C., Chang, S. S., Ostrow, K. L., Westra, W. H., Begum, S., Dhara, M. & Califano, J. 2009, 'Coordinated activation of candidate proto-oncogenes and cancer testes antigens via promoter demethylation in head and neck cancer and lung cancer', *PLoS One*, vol. 4, no. 3, p. e4961.

Smith, M., Arthur, D., Camitta, B., Carroll, A. J., Crist, W., Gaynon, P., Gelber, R., Heerema, N., Korn, E. L., Link, M., Murphy, S., Pui, C. H., Pullen, J., Reamon, G., Sallan, S. E., Sather, H., Shuster, J., Simon, R., Trigg, M., Tubergen, D., Uckun, F. & Ungerleider, R. 1996, 'Uniform approach to risk classification and treatment assignment for children with acute lymphoblastic leukemia', *J Clin Oncol*, vol. 14, no. 1, pp. 18-24.

Snowden, A. W., Gregory, P. D., Case, C. C. & Pabo, C. O. 2002, 'Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo', *Curr Biol*, vol. 12, no. 24, pp. 2159-66.

Somervaille, T. C. & Cleary, M. L. 2010, 'Grist for the MLL: how do MLL oncogenic fusion proteins generate leukemia stem cells?' *Int J Hematol*, vol. 91, no. 5, pp. 735-41.

Song, J. J., Aswad, R., Kanaan, R. A., Rico, M. C., Owen, T. A., Barbe, M. F., Safadi, F. F. & Popoff, S. N. 2007, 'Connective tissue growth factor (CTGF) acts as a downstream mediator of TGF-beta1 to induce mesenchymal cell condensation', *J Cell Physiol*, vol. 210, no. 2, pp. 398-410.

Soreq, L., Gilboa-Geffen, A., Berrih-Aknin, S., Lacoste, P., Darvasi, A., Soreq, E., Bergman, H. & Soreq, H. 2008, 'Identifying alternative hyper-splicing signatures in MG-thymoma by exon arrays', *PLoS One*, vol. 3, no. 6, p. e2392.

Stark, B., Avrahami, G., Nirel, R., Abramov, A., Attias, D., Ballin, A., Bielorai, B., Burstein, Y., Gavriel, H., Elhasid, R., Kapelushnik, J., Sthoeger, D., Toren, A., Wientraub, M., Yaniv, I. & Izraeli, S. 2009, 'Extended triple intrathecal therapy in

children with T-cell acute lymphoblastic leukaemia: a report from the Israeli National ALL-Studies', *Br J Haematol*, vol. 147, no. 1, pp. 113-24.

Steffen, C. L., Ball-Mirth, D. K., Harding, P. A., Bhattacharyya, N., Pillai, S. & Brigstock, D. R. 1998, 'Characterization of cell-associated and soluble forms of connective tissue growth factor (CTGF) produced by fibroblast cells in vitro', *Growth Factors*, vol. 15, no. 3, pp. 199-213.

Stielow, B., Kruger, I., Diezko, R., Finkernagel, F., Gillemans, N., Kong-a-San, J., Philipsen, S. & Suske, G. 2010, 'Epigenetic silencing of spermatocyte-specific and neuronal genes by SUMO modification of the transcription factor Sp3', *PLoS Genet*, vol. 6, no. 11, p. e1001203.

Stielow, B., Sapetschnig, A., Wink, C., Kruger, I. & Suske, G. 2008, 'SUMO-modified Sp3 represses transcription by provoking local heterochromatic gene silencing', *EMBO Rep*, vol. 9, no. 9, pp. 899-906.

Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. 2006, 'Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches', *Immunity*, vol. 25, no. 6, pp. 977-88.

Sun, L., Liu, A. & Georgopoulos, K. 1996, 'Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development', *EMBO J*, vol. 15, no. 19, pp. 5358-69.

Sutcliffe, M. J., Shuster, J. J., Sather, H. N., Camitta, B. M., Pullen, J., Schultz, K. R., Borowitz, M. J., Gaynon, P. S., Carroll, A. J. & Heerema, N. A. 2005, 'High concordance from independent studies by the Children's Cancer Group (CCG) and Pediatric Oncology Group (POG) associating favorable prognosis with combined trisomies 4, 10, and 17 in children with NCI Standard-Risk B-precursor Acute Lymphoblastic Leukemia: a Children's Oncology Group (COG) initiative', *Leukemia*, vol. 19, no. 5, pp. 734-40.

Suzuma, K., Naruse, K., Suzuma, I., Takahara, N., Ueki, K., Aiello, L. P. & King, G. L. 2000, 'Vascular endothelial growth factor induces expression of connective tissue growth factor via KDR, Flt1, and phosphatidylinositol 3-kinase-akt-dependent pathways in retinal vascular cells', *J Biol Chem*, vol. 275, no. 52, pp. 40725-31.

Tabayashi, T., Ishimaru, F., Takata, M., Kataoka, I., Nakase, K., Kozuka, T. & Tanimoto, M. 2007, 'Characterization of the short isoform of Helios overexpressed in patients with T-cell malignancies', *Cancer Sci*, vol. 98, no. 2, pp. 182-8.

Takai, D. & Jones, P. A. 2003, 'The CpG island searcher: a new WWW resource', *In Silico Biol*, vol. 3, no. 3, pp. 235-40.

Takigawa, M., Nakanishi, T., Kubota, S. & Nishida, T. 2003, 'Role of CTGF/HCS24/ecogenin in skeletal growth control', *J Cell Physiol*, vol. 194, no. 3, pp. 256-66.

Tan, T. W., Lai, C. H., Huang, C. Y., Yang, W. H., Chen, H. T., Hsu, H. C., Fong, Y. C. & Tang, C. H. 2009, 'CTGF enhances migration and MMP-13 up-regulation via alphavbeta3 integrin, FAK, ERK, and NF-kappaB-dependent pathway in human chondrosarcoma cells', *J Cell Biochem*, vol. 107, no. 2, pp. 345-56.

Tanaka, M., Oikawa, K., Takanashi, M., Kudo, M., Ohyashiki, J., Ohyashiki, K. & Kuroda, M. 2009, 'Down-regulation of miR-92 in human plasma is a novel marker for acute leukemia patients', *PLoS One*, vol. 4, no. 5, p. e5532.

Tanaka, S., Sugimachi, K., Saeki, H., Kinoshita, J., Ohga, T., Shimada, M., Maehara, Y. & Sugimachi, K. 2001, 'A novel variant of WISP1 lacking a Von Willebrand type C module overexpressed in scirrhous gastric carcinoma', *Oncogene*, vol. 20, no. 39, pp. 5525-32.

Thelander, E. F., Ichimura, K., Corcoran, M., Barbany, G., Nordgren, A., Heyman, M., Berglund, M., Mungall, A., Rosenquist, R., Collins, V. P., Grandner, D., Larsson, C. & Lagercrantz, S. 2008, 'Characterization of 6q deletions in mature B cell lymphomas and childhood acute lymphoblastic leukemia', *Leuk Lymphoma*, vol. 49, no. 3, pp. 477-87.

Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., Yi, E. C., Dai, H., Thorsson, V., Eng, J., Goodlett, D., Berger, J. P., Gunter, B., Linseley, P. S., Stoughton, R. B., Aebersold, R., Collins, S. J., Hanlon, W. A. & Hood, L. E. 2004, 'Integrated genomic and proteomic analyses of gene expression in Mammalian cells', *Mol Cell Proteomics*, vol. 3, no. 10, pp. 960-9.

Tonelli, R., Sartini, R., Fronza, R., Freccero, F., Franzoni, M., Dongiovanni, D., Ballarini, M., Ferrari, S., D'Apolito, M., Di Cola, G., Capranico, G., Khobta, A., Campanini, R., Paolucci, P., Minucci, S. & Pession, A. 2006, 'G1 cell-cycle arrest and apoptosis by histone deacetylase inhibition in MLL-AF9 acute myeloid leukemia cells is p21 dependent and MLL-AF9 independent', *Leukemia*, vol. 20, no. 7, pp. 1307-10.

Trojanowska, M. 2009, 'Noncanonical transforming growth factor beta signaling in scleroderma fibrosis', *Curr Opin Rheumatol*, vol. 21, no. 6, pp. 623-9.

Tsujiuchi, T., Tsutsumi, M., Sasaki, Y., Takahama, M. & Konishi, Y. 1999, 'Hypomethylation of CpG sites and c-myc gene overexpression in hepatocellular carcinomas, but not hyperplastic nodules, induced by a choline-deficient L-amino acid-defined diet in rats', *Jpn J Cancer Res*, vol. 90, no. 9, pp. 909-13.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P. & Zhang, Y. 2006, 'Histone demethylation by a family of JmjC domain-containing proteins', *Nature*, vol. 439, no. 7078, pp. 811-6.

Uckun, F. M., Nachman, J. B., Sather, H. N., Sensel, M. G., Kraft, P., Steinherz, P. G., Lange, B., Hutchinson, R., Reaman, G. H., Gaynon, P. S. & Heerema, N. A. 1998a, 'Clinical significance of Philadelphia chromosome positive pediatric acute lymphoblastic leukemia in the context of contemporary intensive therapies: a report from the Children's Cancer Group', *Cancer*, vol. 83, no. 9, pp. 2030-9.

Uckun, F. M., Sensel, M. G., Sather, H. N., Gaynon, P. S., Arthur, D. C., Lange, B. J., Steinherz, P. G., Kraft, P., Hutchinson, R., Nachman, J. B., Reaman, G. H. & Heerema, N. A. 1998b, 'Clinical significance of translocation t(1;19) in childhood acute lymphoblastic leukemia in the context of contemporary therapies: a report from the Children's Cancer Group', *J Clin Oncol*, vol. 16, no. 2, pp. 527-35.

Ueki, T., Walter, K. M., Skinner, H., Jaffee, E., Hruban, R. H. & Goggins, M. 2002, 'Aberrant CpG island methylation in cancer cell lines arises in the primary cancers from which they were derived', *Oncogene*, vol. 21, no. 13, pp. 2114-7.

Van Beek, J. P., Kennedy, L., Rockel, J. S., Bernier, S. M. & Leask, A. 2006, 'The induction of CCN2 by TGFbeta1 involves Ets-1', *Arthritis Res Ther*, vol. 8, no. 2, p. R36.

van der Velden, V. H. & van Dongen, J. J. 2009, 'MRD detection in acute lymphoblastic leukemia patients using Ig/TCR gene rearrangements as targets for real-time quantitative PCR', *Methods Mol Biol*, vol. 538, pp. 115-50.

Van Dijk, M. A., Voorhoeve, P. M. & Murre, C. 1993, 'Pbx1 is converted into a transcriptional activator upon acquiring the N-terminal region of E2A in pre-B-cell acute lymphoblastoid leukemia', *Proc Natl Acad Sci U S A*, vol. 90, no. 13, pp. 6061-5.

Vaux, D. L., Cory, S. & Adams, J. M. 1988, 'Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells', *Nature*, vol. 335, no. 6189, pp. 440-2.

Volinia, S., Galasso, M., Costinean, S., Tagliavini, L., Gamberoni, G., Drusco, A., Marchesini, J., Mascellani, N., Sana, M. E., Abu Jarour, R., Desponts, C., Teitell, M., Baffa, R., Aqeilan, R., Iorio, M. V., Taccioli, C., Garzon, R., Di Leva, G., Fabbri, M., Catozzi, M., Previati, M., Ambs, S., Palumbo, T., Garofalo, M., Veronese, A., Bottoni, A., Gasparini, P., Harris, C. C., Visone, R., Pekarsky, Y., de la Chapelle, A., Bloomston, M., Dillhoff, M., Rassenti, L. Z., Kipps, T. J., Huebner, K., Pichiorri, F., Lenze, D., Cairo, S., Buendia, M. A., Pineau, P., Dejean, A., Zanesi, N., Rossi, S., Calin, G. A., Liu, C. G., Palatini, J., Negrini, M., Vecchione, A., Rosenberg, A. & Croce, C. M. 2010, 'Reprogramming of miRNA networks in cancer and leukemia', *Genome Res*, vol. 20, no. 5, pp. 589-99.

Vorwerk, P., Wex, H., Hohmann, B., Oh, Y., Rosenfeld, R. G. & Mittler, U. 2000, 'CTGF (IGFBP-rP2) is specifically expressed in malignant lymphoblasts of patients with acute lymphoblastic leukaemia (ALL)', *Br J Cancer*, vol. 83, no. 6, pp. 756-60.

Vucic, E. A., Brown, C. J. & Lam, W. L. 2008, 'Epigenetics of cancer progression', *Pharmacogenomics*, vol. 9, no. 2, pp. 215-34.

Wagner, W., Wein, F., Roderburg, C., Saffrich, R., Faber, A., Krause, U., Schubert, M., Benes, V., Eckstein, V., Maul, H. & Ho, A. D. 2007, 'Adhesion of hematopoietic progenitor cells to human mesenchymal stem cells as a model for cell-cell interaction', *Exp Hematol*, vol. 35, no. 2, pp. 314-25.

Wahab, N. A., Weston, B. S. & Mason, R. M. 2005, 'Connective tissue growth factor CCN2 interacts with and activates the tyrosine kinase receptor TrkA', *J Am Soc Nephrol*, vol. 16, no. 2, pp. 340-51.

Wahlfors, J., Hiltunen, H., Heinonen, K., Hamalainen, E., Alhonen, L. & Janne, J. 1992, 'Genomic hypomethylation in human chronic lymphocytic leukemia', *Blood*, vol. 80, no. 8, pp. 2074-80.

Wallis, J. P. & Reid, M. M. 1989, 'Bone marrow fibrosis in childhood acute lymphoblastic leukaemia', *J Clin Pathol*, vol. 42, no. 12, pp. 1253-4.

Wang, L. C., Swat, W., Fujiwara, Y., Davidson, L., Visvader, J., Kuo, F., Alt, F. W., Gilliland, D. G., Golub, T. R. & Orkin, S. H. 1998, 'The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow', *Genes Dev*, vol. 12, no. 15, pp. 2392-402.

Wang, Q., Usinger, W., Nichols, B., Gray, J., Xu, L., Seeley, T. W., Brenner, M., Guo, G., Zhang, W., Oliver, N., Lin, A. & Yeowell, D. 2011, 'Cooperative interaction of CTGF and TGF-beta in animal models of fibrotic disease', *Fibrogenesis Tissue Repair*, vol. 4, no. 1, p. 4.

Wang, Q., Williamson, M., Bott, S., Brookman-Amissah, N., Freeman, A., Nariculam, J., Hubank, M. J., Ahmed, A. & Masters, J. R. 2007, 'Hypomethylation of WNT5A, CRIP1 and S100P in prostate cancer', *Oncogene*, vol. 26, no. 45, pp. 6560-5.

Wang, Q., Zhang, A., Li, R., Liu, J., Xie, J., Deng, A., Feng, Y. & Zhu, Z. 2008, 'High glucose promotes the CTGF expression in human mesangial cells via serum and glucocorticoid-induced kinase 1 pathway', *J Huazhong Univ Sci Technolog Med Sci*, vol. 28, no. 5, pp. 508-12.

Watanabe, Y. & Maekawa, M. 2010, 'Methylation of DNA in cancer', *Adv Clin Chem*, vol. 52, pp. 145-67.

Weisberg, E., Wright, R. D., McMillin, D. W., Mitsiades, C., Ray, A., Barrett, R., Adamia, S., Stone, R., Galinsky, I., Kung, A. L. & Griffin, J. D. 2008, 'Stromal-mediated protection of tyrosine kinase inhibitor-treated BCR-ABL-expressing leukemia cells', *Mol Cancer Ther*, vol. 7, no. 5, pp. 1121-9.

Welner, R. S., Pelayo, R. & Kincade, P. W. 2008, 'Evolving views on the genealogy of B cells', *Nat Rev Immunol*, vol. 8, no. 2, pp. 95-106.

Wenger, C., Ellenrieder, V., Alber, B., Lacher, U., Menke, A., Hameister, H., Wilda, M., Iwamura, T., Beger, H. G., Adler, G. & Gress, T. M. 1999, 'Expression and differential regulation of connective tissue growth factor in pancreatic cancer cells', *Oncogene*, vol. 18, no. 4, pp. 1073-80.

Whitlock, J. A., Sather, H. N., Gaynon, P., Robison, L. L., Wells, R. J., Trigg, M., Heerema, N. A. & Bhatia, S. 2005, 'Clinical characteristics and outcome of children with Down syndrome and acute lymphoblastic leukemia: a Children's Cancer Group study', *Blood*, vol. 106, no. 13, pp. 4043-9.

Whitman, S. P., Strout, M. P., Marcucci, G., Freud, A. G., Culley, L. L., Zeleznik-Le, N. J., Mrozek, K., Theil, K. S., Kees, U. R., Bloomfield, C. D. & Caligiuri, M. A. 2001, 'The partial nontandem duplication of the MLL (ALL1) gene is a novel rearrangement that generates three distinct fusion transcripts in B-cell acute lymphoblastic leukemia', *Cancer Res*, vol. 61, no. 1, pp. 59-63.

Wiemels, J., Kang, M. & Greaves, M. 2009, 'Backtracking of leukemic clones to birth', *Methods Mol Biol*, vol. 538, pp. 7-27.

Wilson, A., Laurenti, E. & Trumpp, A. 2009, 'Balancing dormant and self-renewing hematopoietic stem cells', *Curr Opin Genet Dev*, vol. 19, no. 5, pp. 461-8.

Wilson, A., Oser, G. M., Jaworski, M., Blanco-Bose, W. E., Laurenti, E., Adolphe, C., Essers, M. A., Macdonald, H. R. & Trumpp, A. 2007, 'Dormant and self-renewing hematopoietic stem cells and their niches', *Ann NY Acad Sci*, vol. 1106, pp. 64-75.

Wilson, A. & Trumpp, A. 2006, 'Bone-marrow haematopoietic-stem-cell niches', *Nat Rev Immunol*, vol. 6, no. 2, pp. 93-106.

Wood, M. A., McMahon, S. B. & Cole, M. D. 2000, 'An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc', *Mol Cell*, vol. 5, no. 2, pp. 321-30.

Xie, S., Sukkar, M. B., Issa, R., Oltmanns, U., Nicholson, A. G. & Chung, K. F. 2005, 'Regulation of TGF-beta 1-induced connective tissue growth factor expression in airway smooth muscle cells', *Am J Physiol Lung Cell Mol Physiol*, vol. 288, no. 1, pp. L68-76.

Xie, Y., Yin, T., Wiegand, W., He, X. C., Miller, D., Stark, D., Perko, K., Alexander, R., Schwartz, J., Grindley, J. C., Park, J., Haug, J. S., Wunderlich, J. P., Li, H., Zhang, S., Johnson, T., Feldman, R. A. & Li, L. 2009, 'Detection of functional haematopoietic stem cell niche using real-time imaging', *Nature*, vol. 457, no. 7225, pp. 97-101.

Xu, Y., Fang, F., Dhar, S. K., Bosch, A., St Clair, W. H., Kasarskis, E. J. & St Clair, D. K. 2008, 'Mutations in the SOD2 promoter reveal a molecular basis for an activating protein 2-dependent dysregulation of manganese superoxide dismutase expression in cancer cells', *Mol Cancer Res*, vol. 6, no. 12, pp. 1881-93.

Yagi, T., Hibi, S., Takanashi, M., Kano, G., Tabata, Y., Imamura, T., Inaba, T., Morimoto, A., Todo, S. & Imashuku, S. 2002, 'High frequency of Ikaros isoform 6 expression in acute myelomonocytic and monocytic leukemias: implications for up-regulation of the antiapoptotic protein Bcl-XL in leukemogenesis', *Blood*, vol. 99, no. 4, pp. 1350-5.

Yanagita, T., Kubota, S., Kawaki, H., Kawata, K., Kondo, S., Takano-Yamamoto, T., Tanaka, S. & Takigawa, M. 2007, 'Expression and physiological role of CCN4/Wnt-induced secreted protein 1 mRNA splicing variants in chondrocytes', *FEBS J* vol. 274, no. 7, pp. 1655-65.

Yang, D. H., Kim, H. S., Wilson, E. M., Rosenfeld, R. G. & Oh, Y. 1998, 'Identification of glycosylated 38-kDa connective tissue growth factor (IGFBP-related protein 2) and proteolytic fragments in human biological fluids, and up-regulation of IGFBP-rP2 expression by TGF-beta in Hs578T human breast cancer cells', *J Clin Endocrinol Metab*, vol. 83, no. 7, pp. 2593-6.

Yang, F., Tuxhorn, J. A., Ressler, S. J., McAlhany, S. J., Dang, T. D. & Rowley, D. R. 2005, 'Stromal expression of connective tissue growth factor promotes angiogenesis and prostate cancer tumorigenesis', *Cancer Res*, vol. 65, no. 19, pp. 8887-95.

Yang, M., Huang, H., Li, J., Li, D. & Wang, H. 2004, 'Tyrosine phosphorylation of the LDL receptor-related protein (LRP) and activation of the ERK pathway are required for connective tissue growth factor to potentiate myofibroblast differentiation', *FASEB J*, vol. 18, no. 15, pp. 1920-1.

Yi, F., Pereira, L. & Merrill, B. J. 2008, 'Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal', *Stem Cells*, vol. 26, no. 8, pp. 1951-60.

Yin, D., Chen, W., O'Kelly, J., Lu, D., Ham, M., Doan, N. B., Xie, D., Wang, C., Vadgama, J., Said, J. W., Black, K. L. & Koeffler, H. P. 2010, 'Connective tissue growth factor associated with oncogenic activities and drug resistance in glioblastoma multiforme', *Int J Cancer*, vol. 127, no. 10, pp. 2257-67.

Yoshida, N., Yagasaki, H., Xu, Y., Matsuda, K., Yoshimi, A., Takahashi, Y., Hama, A., Nishio, N., Muramatsu, H., Watanabe, N., Matsumoto, K., Kato, K., Ueyama, J., Inada, H., Goto, H., Yabe, M., Kudo, K., Mimaya, J., Kikuchi, A., Manabe, A., Koike, K. & Kojima, S. 2009, 'Correlation of clinical features with the mutational status of GM-CSF signaling pathway-related genes in juvenile myelomonocytic leukemia', *Pediatr Res*, vol. 65, no. 3, pp. 334-40.

Youlden, D., Baade, P., Ward, L., Valery, P., Hassell, T., Green, A. C. & Aitken, J. F. 2010, *Childhood cancer survival in Australia, 1995-2004*, Viertel Center for Research in Cancer Control, Brisbane, Queensland. Accessed: Jan 2010.

Yu, C. C., Hsu, M. J., Kuo, M. L., Chen, R. F., Chen, M. C., Bai, K. J., Yu, M. C., Chen, B. C. & Lin, C. H. 2009, 'Thrombin-induced connective tissue growth factor expression in human lung fibroblasts requires the ASK1/JNK/AP-1 pathway', *J Immunol*, vol. 182, no. 12, pp. 7916-27.

Zhang, Q., Ying, J., Zhang, K., Li, H., Ng, K. M., Zhao, Y., He, Q., Yang, X., Xin, D., Liao, S. K., Tao, Q. & Jin, J. 2007, 'Aberrant methylation of the 8p22 tumor suppressor gene DLC1 in renal cell carcinoma', *Cancer Lett*, vol. 249, no. 2, pp. 220-6.

Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., Yu, J., Lin, J. D., Wang, C. Y., Chinnaiyan, A. M., Lai, Z. C. & Guan, K. L. 2008, 'TEAD mediates YAP-dependent gene induction and growth control', *Genes Dev*, vol. 22, no. 14, pp. 1962-71.

Zivny, J., Klener, P., Jr., Pytlik, R. & Andera, L. 2010, 'The role of apoptosis in cancer development and treatment: focusing on the development and treatment of hematologic malignancies', *Curr Pharm Des*, vol. 16, no. 1, pp. 11-33.

Zumkeller, W. 2002, 'The insulin-like growth factor system in hematopoietic cells', *Leuk Lymphoma*, vol. 43, no. 3, pp. 487-91.

Chapter 1

General Introduction

Chapter 1

General Introduction

1.1 Childhood cancer

In Australia, cancer is the leading cause of disease-associated deaths among children aged 1-14 years (Australian Bureau of Statistics 2006). This statistic is universal in the developed world, with childhood cancer outstripping deaths associated with other disease processes in children (Kaatsch 2010). While outcomes for some childhood cancers today have improved dramatically, there is still much work to be done to improve our understanding of underlying cancer biology. By enhancing our knowledge of the molecular mechanisms involved in oncogenesis we can continue to develop new and refine existing therapies to improve clinical outcomes.

Of the various types of cancers commonly affecting infants (< 1yr) and children (1-14 yrs), leukaemia is by far the most prevalent accounting for around 34% of new diagnoses, followed by cancers of the central nervous system at around 24% (Pieters & Carroll 2010). Leukaemia is a malignancy arising from transformation of cells of the haemopoietic system and there are two main types, lymphoblastic and myelogenous, owing to the two arms of haemopoiesis, myeloid and lymphoid which both arise from a common stem or progenitor cell. Acute lymphoblastic leukaemia (ALL) is the most common form of the disease in children accounting for 79% of all paediatric leukaemia diagnoses. This equates to 27% of all childhood cancers (Kaatsch 2010), thus, leukaemia represents a significant contribution to childhood mortality and morbidity. Improving our understanding of the molecular mechanisms of leukaemogenesis is critical to improve outcomes for young patients and their families.

1.2 Acute lymphoblastic leukaemia

As the name suggests, ALL is a swiftly progressive form of leukaemia, characterised by the clonal accumulation of rapidly proliferating, poorly differentiated lymphoid

precursors (Bourquin & Izraeli 2010). ALL disease can be divided into two broad biological categories, precursor B-cell ALL (pre-B ALL) and precursor T-cell ALL (T-ALL) (Gaynon 2005). Of these two types, pre-B ALL is more common, accounting for up to 85% of childhood ALL diagnoses (Pieters & Carroll 2010). Presenting symptoms invariably correlate with the level of tumour cell burden within the bone marrow. As normal marrow is replaced by leukaemic cells, haemopoiesis is impaired, resulting in blood lineage cytopenias and if left untreated ALL is rapidly fatal (Onciu 2009). Other common symptoms include fever, fatigue and joint or bone pain. Leukaemic cells can also reside outside of the bone marrow and common sites of extramedullary involvement include the blood, liver, spleen, lymph nodes, meninges and in T-lineage ALLs, the thymus.

1.2.1 Origins: Normal haemopoiesis

The haemopoietic system is responsible for the production of differentiated cells of all blood lineages. This includes lymphoid cells, incorporating B, T and NK cells, as well as those of the myeloid lineage, such as erythrocytes, megakaryocytes and granulocytes. All of these mature blood cells are formed from a common haemopoietic stem cell (HSC) at the apex of this developmental hierarchy (Mansson et al. 2009). These cells reside in the bone marrow space of long bones in specialised niches that are tightly regulated (Wilson & Trumpp 2006). HSCs are normally quiescent, but will respond to haemopoietic stress such as blood loss by re-entering the cell cycle and dividing to produce intermediate progenitor cells which in turn give rise to mature functional effector cells. Haemopoiesis progresses by limiting the developmental potential of precursors in a step-wise fashion until the fully mature blood cells are produced (Blom & Spits 2006). This process has been extensively characterised in mice and while surface antigens differ between species, the overall process of differentiation from HSC to mature blood cells is thought to be highly conserved (Welner et al. 2008; Mansson et al. 2009).

1.2.1.1 Haemopoietic stem cell niche

The bone marrow microenvironment is comprised of a number of different cell types, including reticular cells, fibroblasts, chondrocytes, adipocytes, endothelial cells, osteoblasts and osteoclasts, as well as haemopoietic stem and progenitor cells (Ehninger & Trumpp 2011). Defining the role of each of these cell types within the context of haemopoiesis is still ongoing, however it is widely accepted that haemopoietic stem cells (HSCs) reside in specialised regions within the bone marrow microenvironment that govern HSC fate by providing extrinsic regulatory signals. These developmental sites, termed niches, provide extracellular cues through soluble factors and cell-cell interactions that are critical for guiding the maintenance, proliferation and differentiation of HSCs throughout adult life (Mendez-Ferrer et al. 2010; Nagasawa 2006; Sugiyama et al. 2006; Wilson et al. 2007).

The mouse haemopoietic stem cell is possibly the most-well characterised of all mammalian stem cells. These cells are well defined by surface marker expression (lin^{neg}Scal^{hi}c-Kit⁺CD34⁻CD48⁻CD150^{hi}), and can reconstitute the entire haemopoietic system of irradiated mice (Purton & Scadden 2007; Wilson et al. 2007). Advances in intravital imaging have revealed the localisation of these HSCs within distinct regions of bone marrow including the endosteal surface (termed the endosteal niche) as well as adjacent to sinusoidal endothelium (termed the vascular niche) (Xie et al. 2009; Lo Celso et al. 2009), and these two niches are thought to differentially regulate stem cell fate. Evidence is now emerging for the central role of mesenchymal stem cells (MSCs) within the haemopoietic niche in maintaining the bone marrow HSC pool. In recent years, a subset of perivascular nestin positive MSCs have been identified that are critical for HSC maintenance (Mendez-Ferrer et al. 2010). These nestin^{pos} cells also express high levels of the HSC maintenance factors chemokine C-X-C motif ligand 12 (Cxcl12), osteopontin, angiopoietin-1, interleukin 7 and vascular cell adhesion molecule 1, and depletion of nestin^{pos} MSCs results in the mobilisation of HSCs away from the bone marrow to other organs (Mendez-Ferrer et al. 2010). Furthermore, the bone marrow-homing of transplanted HSCs into nestin^{pos} MSC-depleted recipients is significantly reduced. Similar to nestin^{pos} MSCs, a distinct subset of reticular cells also express high

levels of Cxcl12 and these cells dubbed Cxcl12 abundant reticulocytes (CAR) are found in close proximity to HSCs within the vascular niche and are critical for lymphoid specification of haematopoietic progenitor cells (Crisan et al. 2008; Omatsu et al. 2010).

1.2.1.2 Lymphoid specification

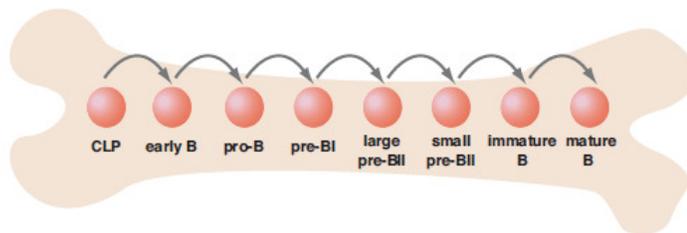
Fluorescence activated cell sorting (FACS) has allowed the prospective isolation and characterisation of cells at various stages of development with particular emphasis on their lineage potential (Mansson et al. 2009). The previously accepted model of haemopoiesis proposed a linear series of maturation steps, beginning with asymmetric division of activated HSCs giving rise to multipotent progenitor cells (MPPs), which in turn gave rise to either common myeloid progenitors (CMPs) or common lymphoid progenitors (CLP). Recent evidence has pointed to several intermediate steps in lymphoid specification prior to the restriction to CLP identity, however the precise model of transition from MPP to CLP in humans is still unclear (Corfe & Paige 2009). The earliest stage of irreversible lymphoid specification of CLPs is marked by expression of the interleukin 7 receptor (*IL7R/CD127*) (Dias et al. 2005; Parrish et al. 2009). These cells can progress to form mature B and T lymphocytes and NK cells.

1.2.1.3 B-cell development

Humoral immunity is critical to mammalian homeostasis and B cells are the effectors of this arm of the immune system. Mature B cells produce antibodies against foreign antigens and can also act as antigen-presenting cells themselves (Rodriguez-Pinto 2005). Activation of the early B cell factor (*EBF*) gene is critical for the induction of B-cell development. *EBF* activates *PAX5* which acts as a positive regulator of *EBF*, enforcing a feedback loop. These two transcription factors then act in concert to inhibit T-lineage and myeloid gene expression programs.

While the model of lymphoid lineage specification is still evolving, the process of B-cell development from CLP to mature B-cell in humans is well established (Blom & Spits

2006). This progression, outlined in Figure 1.1, is marked by a number of immunophenotypic changes at each stage of development. These include differential expression of cluster of differentiation (CD) markers; CD34, CD10, CD127 (IL7R α), CD19, CD79a, and CD179a (Vpre-B) (Blom & Spits 2006). As cells progress through B-cell development, they can also be characterized by the presence or absence of specific immunoglobulin (*Ig*) gene rearrangements, which are important events in B lymphopoiesis. Expression of terminal deoxynucleotidyl transferase (*TdT*) and recombination activating gene (*RAG*) are both required for successful rearrangement of *Ig* loci and are expressed early in B-cell development (Corfe & Paige 2009). Prior to the immature B stage, development occurs exclusively within the bone marrow. Acquisition of surface IgM expression marks the stage at which immature B-cells exit the bone marrow and home to the spleen and other secondary lymphoid organs where they develop into mature B-cells.



CD34	+	+	+	-	-	-	-	-
CD10	+	+	+	+	+	+	+	-
IL-7R α	+	+	+	-	-	-	-	-
CD19	-	-	+	+	+	+	+	+
CD79a	-	+	+	+	+	+	+	+
TdT	-	-	+	-	-	-	-	-
RAG	-	-	+	+	-	+	+	-
Vpre-B	-	+	+	+	+	-	-	-
μ H	-	-	+/-	+	+	+	+	+
pre-BCR	-	-	-	-	+	-	-	-
IgH	GL	DJ _H	V _H DJ _H					
κ L	GL	GL	GL	GL	GL	V _L J _L	V _L J _L	V _L J _L
cycling	-	-	-	+	+	-	-	-
Pax-5	-	-	+	+	+	+	+	+
sIgM	-	-	-	-	-	-	+	+
sIgD	-	-	-	-	-	-	-	+

Figure 1.1 Model of human B-cell development

Adapted from Blom and Spits (2006).

1.2.2 Origins: Aetiology and incidence

Pre-B ALL has a peak incidence at 2-5 years of age (Gurney et al. 1995) and approximately one hundred and sixty new cases of ALL are diagnosed annually in Australia, with this number remaining unchanged over the last decade (Youlten et al. 2010). There is evidence that acquisition of genetic changes associated with ALL can occur *in utero* (Maia et al. 2003; Hjalgrim et al. 2002). This is supported by retrospective detection of genetic signatures associated with some forms of childhood ALL in Guthrie spot cards taken at birth from children who later developed ALL (Wiemels et al. 2009; Hjalgrim et al. 2002). Recurrent genetic abnormalities (discussed in detail later) have been associated with increased risk of developing ALL in early life and children with Down's syndrome (trisomy 21) are 10-15 times more likely to develop ALL, and they often suffer more treatment-related toxicities (Whitlock et al. 2005). Race and ethnicity has been shown to predict significantly different outcomes for children diagnosed with ALL, with children of African or Hispanic ancestry having a poorer prognosis than those with Anglo-Saxon ancestry (Bhatia 2004; McNeil et al. 2002), although this has been attributed by some to the lack of parity in healthcare for some minorities (Pui et al. 2003; Pui & Evans 2006). There is also a predilection towards a higher incidence of ALL in males, however the reason for this bias is unclear. *In utero* exposure to solvents, pesticides, hydrocarbons or ionizing radiation have all been linked to increased risk of developing ALL in early life (McNally & Parker 2006; Buffler et al. 2005), and studies on this topic are ongoing.

1.2.3 ALL therapy

Cancer in the very young can lead to significant problems for survivors in later life (Dowling et al. 2010). Cancer treatments in early life may result in impaired neurocognitive development and function (Ness et al. 2010), infertility, cardiovascular disease (Geenen et al. 2010), as well as impairments in musculoskeletal development (Ishida et al. 2010) and other complications. Furthermore, the experiences of cancer survivors can result in significant psychosocial disorders (Oeffinger et al. 2008). While current therapeutic regimens are largely successful, the present goal is to develop

therapies that are less toxic and will ultimately result in improved long term outcomes for survivors. Reducing the severity of treatment, while still attaining curative outcomes is important, thus balancing the type and amount of therapy with the risk of relapse is critical to reduce unnecessary exposure to toxic agents.

Despite the relatively high incidence of ALL in children, 5 yr event free survival (EFS) of up to 85-90% can now be achieved for some forms of the disease (Conter et al. 2010; Gaynon et al. 2010; Pui et al. 2010). Improved survival rates are attributable to the optimisation of complex chemotherapeutic regimens assessed in large scale, randomized clinical trials, as well as improved understanding of ALL biology (Pieters & Carroll 2010). Risk-based stratification of patients has also resulted in improved patient outcomes, and patients that are recognized as having a high risk of relapse are now receiving more intensive therapy. Unfortunately, those patients that do relapse often have a dismal prognosis, with relapsed ALL usually resistant to front-line chemotherapy.

Clinical management of ALL is continually undergoing improvements through large scale multicenter clinical trials run by major cooperative groups such as the Children's Oncology Group (COG) and the United Kingdom Children's Cancer Study Group (UKCCSG). These groups advise on best-practice treatment for a range of childhood cancers. Therapeutic protocols for the treatment of ALL endorsed by these organisations are similar in that they are comprised of several discreet stages, including induction therapy, consolidation, re-induction and maintenance therapy. However factors such as dosage, timing and the combination of drugs given to patients, differs among groups.

1.2.3.1 Induction therapy

Treatment for ALL incorporates several stages and in most cases continues for up to three years. The first stage of treatment is induction therapy and the primary goal of this phase is to restore normal haemopoiesis by eliminating as many leukaemic blasts as possible with high-dose chemotherapy. Three to four drugs are used depending on the perceived risk classification of patients (discussed later). Standard risk patients will

typically receive vincristine and asparaginase, as well as a glucocorticoid, either dexamethasone, prednisolone or prednisone (McNeer & Nachman 2010). Those patients that are identified as high risk may also receive an anthracycline, especially if they are over the age of 10 years (Pieters & Carroll 2010). The aim of induction chemotherapy is to achieve morphological remission in 4-6 weeks, and a patient's response to induction therapy has invariably been found to be the best predictor of long-term outcome.

The efficacy of induction therapy is evaluated by the detection of minimal residual disease (MRD) which can be defined as the level of leukaemic cells detectable in the bone marrow upon completion of induction therapy. MRD is used as a key indicator of response that guides the remainder of therapy during intensification and consolidation. The two main methods of MRD measurement are quantitative polymerase chain reaction (qPCR) and flow cytometry. The current limit of detection by flow cytometry is between 1 cell in 10^3 to 10^4 , while assessment of immunoglobulin (*IG*) or T-cell receptor (*TCR*) rearrangements by qPCR affords sensitivity of 1 cell in 10^4 to 10^5 (Bruggemann et al. 2010). While qPCR offers a superior level of sensitivity, this approach requires characterisation of rearrangements that are specific to the patients' leukaemic cells by sequencing the rearranged *IG* and *TCR* loci at diagnosis to develop specific primer/probe combinations. Leukaemic clones can continue to rearrange these loci during the progression of the disease, resulting in false-negatives which may inaccurately guide the remainder of therapy (Bruggemann et al. 2010; van der Velden & van Dongen 2009). Assessing a negative qPCR-MRD finding by flow cytometry offers confirmation, despite the obvious disparity in sensitivity between techniques.

1.2.3.2 Consolidation and re-induction

Once a negative MRD finding indicates that induction therapy has been effective in reducing the burden of leukaemic cells, patients begin the consolidation phase of treatment. Consolidation therapy aims to eliminate residual leukaemic cells that may be cycling slowly or have been driven into growth-arrest in response to chemotherapy. This phase typically includes asparaginase and high-dose methotrexate or 6-mercaptopurine.

Consolidation may also include intrathecal administration (injection into the cerebro-spinal fluid) of methotrexate with or without cytarabine and hydrocortisone to prevent central nervous system (CNS) relapse, particularly in high-risk patients (Pui & Evans 2006). Intrathecal administration of chemotherapy has been shown to have an equal or greater efficacy in preventing CNS relapse compared to cranial irradiation used by some groups (Stark et al. 2009), and introduction of intrathecal CNS prophylaxis has seen a substantial decrease in the practice of cranial radiation, which often has devastating long term effects on the developing brain. Upon completion of consolidation therapy, patients undergo re-induction therapy, sometimes referred to as delayed intensification therapy. This is essentially a repeat of the same therapy used in the induction phase and is a critical component of successful ALL treatment protocols (Pui & Evans 2006).

1.2.3.3 Maintenance therapy

At the completion of re-induction therapy, patients progress to maintenance therapy, which is the final stage of treatment. Maintenance therapy in paediatric ALL can continue for up to two years in females and three years in males. Standard maintenance therapy consists of daily 6-mercaptopurine and weekly methotrexate (Pieters & Carroll 2010). Some study groups have proposed the incorporation of pulsed administration of vincristine and dexamethasone, however the results of trials so far have not shown significant improvements in patient outcomes (Nagatoshi et al. 2010; Arico et al. 2008).

1.3 B-lineage acute lymphoblastic leukaemia

Although traditionally grouped together, cases of B-lineage ALL can be further delineated into different subtypes based on the immunophenotype of leukaemic cells. These subtypes are associated with differentiation stages of B-cell development and have significant prognostic ramifications (Pieters & Carroll 2010; Bene 2005). At the time of diagnosis, patient samples are assessed for immunophenotype and are grouped into one of three types; early pro-B, pre-B and transitional pre-B, while mature B-cell ALL is rarely diagnosed in children (Pui 2006). Table 1.1 outlines these subtypes, and

their immunological features and frequency. It is important to note that there is a degree of plasticity in the immunophenotype of B-lineage ALL cells, making a determination as to the specific subtype difficult in some cases. For this reason, B-lineage ALL is frequently referred to as precursor B-cell ALL (pre-B ALL) as a whole, and often more importance is placed on identifying specific genomic alterations associated with clinical outcomes, as these are considered more important for risk-based stratification.

Table 1.1 Immunologic classification of B-lineage ALL

Table adapted from Childhood Leukemias by Ching-Hon Pui (Pui 2006).

Subtype	Immunological marker (% pos)								frequency
	CD19	CD20	CD22	CD79a	CD10	cIgμ	sIgμ	sIg κ or λ	
Early pre-B (pro-B)	100	35	99	99	95	0	0	0	60-65%
Pre-B	100	45	100	100	100	100	0	0	20-25%
Transitional pre-B (late pre-B)	100	55	100	100	100	100	100	0	10-12%
Mature B	100	99	100	99	50	100	~95	~95	3-5%

Abbreviations: cIgμ, cytoplasmic immunoglobulin mu heavy chain; sIgμ, surface immunoglobulin mu heavy chain; sIgκ or λ, surface immunoglobulin kappa or lambda light chain; CD, cluster of differentiation.

1.3.1 Prognosis and risk stratification

As stated previously, vast improvements have been achieved in the long term EFS of paediatric ALL patients. A cornerstone of contemporary ALL management is the risk-based stratification of patients based on presenting features (Pui et al. 2011). By directing patients expected to have a poor outcome to more intensive therapy, these high-risk patients are afforded the best possible chance of disease clearance. Furthermore, by limiting the amount of therapy given to patients identified as having a favourable outcome, these low-risk patients are spared the detrimental short and long-term side effects associated with agents used in high-risk therapy (Pui & Evans 2006).

Clinical features that predict outcome include age, gender and white blood cell (WBC) count at diagnosis. Biological factors with prognostic relevance include immunophenotype (pro-B, pre-B or common B) and the genotype of the patient's

leukaemic cells. Hyperdiploidy defined as > 51 chromosomes or a DNA index greater than 1.16 is associated with a favourable outcome in standard risk pre-B ALL (Heerema et al. 2000; Sutcliffe et al. 2005). Several recurrent cytogenetic abnormalities are associated with ALL biology and are discussed in the next section. Table 1.2 shows the clinical and biological features that are used to initially stratify patients diagnosed with pre-B ALL as either standard or high-risk. Features associated with a higher risk include age of 0-1yr or greater than 9yrs, male gender, high WBC count ($> 50 \times 10^9/L$), pro-B immunophenotype and hypodiploidy defined as less than 45 chromosomes. As discussed earlier, perhaps the most important prognostic factor in ALL is the response to induction therapy detected as MRD. This is a strong predictor of relapse or resistance to therapy and determines the appropriate treatment protocol for patients, either intensification or consolidation of treatment administered.

Table 1.2 Clinical and biological features for initial patient stratification

Adapted from Pieters and Carroll (2010).

Factor	Standard Risk	High Risk
Age at diagnosis	1-9 yrs	< 1yr or > 9 yrs
Sex	Female	Male
WBC count	$< 50 \times 10^9 / L$	$> 50 \times 10^9 / L$
Immunophenotype	precursor B ALL	B ALL
Genotype	Hyperdiploid >50 chromosomes	hypodiploidy < 45 chromosomes
Post-induction MRD	no blasts detected	detectable blasts

1.3.2 Molecular types and cytogenetics

Numerous recurrent cytogenetic anomalies have been characterised in pre-B ALL and these are associated with different outcomes (Pui & Evans 2006). Fusion transcripts generated by some chromosomal rearrangements have been demonstrated to have leukaemogenic potential. Juxtaposition of genes encoding developmentally critical DNA-binding proteins with foreign loci can result in the production of fusion proteins

with unrestricted and, or, oncogenic transcriptional activity (Poppe et al. 2007). A high proportion of cytogenetic rearrangements associated with leukaemia involve transcription factors required for, or involved in, lymphoid development. Some examples of these are the master transcriptional regulator mixed lineage leukaemia (*MLL*) or the runt-related transcription factor 1 (*RUNX1* also known as *AML1*). The most common cytogenetic alterations and their prognostic implications are outlined in Table 1.3.

1.3.2.1 *TEL-AML1*

The *TEL-AML1* fusion (also known as *ETV6-RUNX1*) is characterised by the t(12;21)(p13;q22) translocation and is the most common cytogenetic rearrangement in paediatric pre-B ALL, occurring in around 25% of cases (Onciu 2009). *TEL-AML1* is associated with high chemosensitivity, and thus a favourable outcome. Both *TEL/ETV6* and *AML1/RUNX1* are transcription factors with established roles in haemopoiesis (Okumura et al. 2007; Wang et al. 1998). Targeted disruption of *TEL* after the establishment of haemopoiesis results in an accumulation of progenitor cells at the pro-B stage, however the developmental potential of these cells is not blocked completely (Fischer et al. 2005). These data reflect what is seen in human disease, as carriers of the *TEL-AML1* rearrangement do not always progress to overt leukaemia and identical twin studies have demonstrated that it is possible to carry *TEL-AML1* in haemopoietic cells without any manifestation of disease (Maia et al. 2003). In studies conducted by Greaves and colleagues, *TEL-AML1*-transduced human CD34^{pos} cord blood cells, transplanted into NOD-SCID mice, gave rise to cells that display B-cell developmental potential and also exhibit greatly enhanced resistance to some apoptotic stimuli, including Fas ligand, camptothecin and melphalan, suggesting that these pre-leukaemic clones have a growth advantage over non-transduced cells (Hong et al. 2008). In carriers of *TEL-AML1*, progenitor cells harbouring this feature are referred to as pre-leukaemic clones and tend to display an early pre-B/pro-B immunophenotype. It is postulated that they require a second hit to become overtly tumorigenic (Greaves 2009).

Table 1.3 Common cytogenetic abnormalities in pre-B ALL

Adapted from Onciu, M. (2009).

Cytogenetic group	Frequency (%)	Fusion gene	Cytogenetic abnormality	Prognosis	Notes
Hyperdiploid	27-29	NA	51-65 chromosomes	Low-risk	Higher sensitivity to MTX, MP
ALL with t(12;21)	25	<i>TEL-AML1</i> (<i>ETVX-RUNX1</i>)	t(12;21)(p13;q22)	Low-risk	Higher sensitivity to L-asparaginase
Hypodiploid	5-6	NA	<45 chromosomes	High-risk	B-cell developmental gene mutation in 100% of cases
ALL with t(1;19)	3-6	<i>TCF3-PBX1</i> (<i>E2A-PBX1</i>)	t(1;19)(q23;p13)	Standard-risk	NA
Philadelphia positive ALL	3-5	<i>BCR-ABL</i>	t(9;22)(q34;q11.2)	High-risk	<i>IKZF1</i> (Ikaros) deletions common
ALL with <i>MLL</i> (11q23) rearranged	2-3	<i>AF4-MLL</i> <i>ELL-MLL</i>	t(4;11)(q21;q23) t(19;11)(p13;q23)	High-risk High-risk	Increased expression of <i>HOX</i> genes.

1.3.2.2 *TCF3-PBX1*

The t(1;19)(q23;p13) *TCF3-PBX1* (also known as *E2A-PBX1*) translocation occurs in around 5% of cases of paediatric B-lineage ALL. Expression of this fusion gene produces an oncogenic protein containing the N-terminal transactivation domain encoded by the transcription factor 3 gene (*TCF3*) fused to the DNA-binding homeodomain encoded by pre-B-cell leukaemia homeobox 1 gene (*PBX1*). The resulting fusion protein can induce expression of genes defined by the *PBX1* DNA-binding domain alone (Van Dijk et al. 1993) or alternatively the *TCF3-PBX1* oncoprotein can dimerize with *HOX* proteins, resulting in enhanced expression of *HOX* target genes (Aspland et al. 2001). *TCF3* encodes two basic helix-loop-helix (bHLH) proteins, E12 and E47 that form homodimeric complexes to regulate immunoglobulin gene expression (Bain et al. 1999). The *TCF3* gene is required for lymphoid development and disruption of normal *TCF3* expression results in accumulation of cells at the pro-B cell stage (Uckun et al. 1998b). Recent studies investigating conditional knockout of *Tcf3* in murine embryonic stem cells have demonstrated that *Tcf3* gene expression plays a critical role in repressing expression of genes associated with self-renewal such as *Nanog*, *Tcl1*, *Tbx3* and *Esrrb* (Yi et al. 2008). Dysregulated expression of the *TCF3-PBX1* may result in improper activation of gene expression programs associated with self-renewal.

The t(1;19) feature was historically associated with a poor outcome, however improvements in multi-agent therapy have resulted in this rearrangement being removed from risk stratification matrices (Raimondi et al. 1990). This may need to be reviewed however, as a recent study has identified an association between *TCF3-PBX1* and CNS relapse in children with pre-B ALL (Jeha et al. 2009). Re-inclusion of this genotype in clinical matrices would direct these patients to intensified CNS-directed prophylaxis which may have a significant effect on patient outcomes.

1.3.2.3 *BCR-ABL1*

Perhaps one of the best-known translocations associated with leukaemia is the t(9;22)(q34;q11.2) *BCR-ABL1* fusion, also known as the Philadelphia chromosome. This rearrangement is characterised by translocation of the coding region of *ABL1* from chromosome 9 to the *BCR* gene on chromosome 22 and occurs in 3-5% of cases of childhood ALL (Arico et al. 2000). This event results in constitutive expression of the *BCR-ABL1* fusion RNA driven by the *BCR* promoter (Pieters & Carroll 2010). While the precise function of the *BCR* gene has not been established, *ABL1* is a proto-oncogene encoding a non-receptor tyrosine kinase and the *BCR-ABL1* fusion protein has been demonstrated as a driver of oncogenesis (Uckun et al. 1998a). Transformation by *BCR-ABL1* activates a number of intracellular signalling pathways, and those shown to play major roles in *BCR-ABL1* induced cellular proliferation and transformation include activation of RAS, PI-3 kinase and Jun kinase (Pui 2006). *BCR-ABL1* transformation also enhances resistance to some cytostatic drugs and radiation by upregulating the anti-apoptotic *BCL2* family member BCL-XL (Skorski 2002). Children with ALL positive for the *BCR-ABL1* genotype usually have a high MRD after induction therapy and this translocation is associated with high-risk and a poor outcome with 5ys EFS of around 28% (Arico et al. 2000).

BCR-ABL1-positive ALL frequently harbours deletions in B-cell developmental genes (Onciu 2009). Recent data gathered on *BCR-ABL1*-positive ALL indicated that 75% of children and 91% of adults harboured mutation of at least one *IKAROS* allele (Mullighan et al. 2008). Iacobucci and colleagues recently reported the acquisition of mutations in *IKAROS* splice sites after treatment with tyrosine kinase inhibitors, suggesting a role for dominant-negative *IKAROS* isoforms in the acquisition of Imatinib and Dasatanib resistance in *BCR-ABL1*-positive patients (Iacobucci et al. 2008).

1.3.2.4 *MLL* rearrangements

Translocations involving the mixed lineage leukaemia (*MLL*) gene at 11q23 are very common in infant ALL (< 1 yr) with around 80% of such cases attributable to

rearrangements of the *MLL* locus. In contrast, only 2% of children aged 1-9 yr present with *MLL* rearrangements (Somerville & Cleary 2010). Universally, *MLL* involvement is associated with high-risk and a poor outcome. Leukaemic cells containing *MLL* rearrangements are usually highly resistant to glucocorticoids and L-asparaginase, two front-line chemotherapeutic agents used in standard induction therapy (Pieters & Carroll 2010). *MLL*-rearranged ALL exhibits an early pro-B immunophenotype and CD10 negativity and mutations in B-cell developmental genes are common (Ayton & Cleary 2001). The striking feature of *MLL* rearrangements is the large number of partner loci involved in 11q23 translocations. To date over 100 different *MLL* rearrangements have been described (Meyer et al. 2009).

MLL is required for normal haemopoiesis (Jude et al. 2007), however the precise mechanism by which *MLL* rearrangements induce transformation is unclear. *MLL* encodes a large, multidomain protein (3972 amino acids), that contains an AT hook DNA-binding domain, as well as a H3K4 histone methyltransferase moiety which is associated with epigenetic activation of homeobox genes by chromatin remodelling. Regardless of the 11q23 fusion partner, *MLL*-fusion oncoproteins consistently induce high level transcriptional activation of *HOXA* and *MEIS1* genes through epigenetic activation of these gene loci (Somerville & Cleary 2010). These genes are normally repressed during terminal differentiation and constitutive expression is thought to contribute to uncontrolled proliferation of leukaemic cells harbouring *MLL* fusions.

1.3.2.5 Numerical chromosomal alterations

Changes in chromosome number (ploidy) are frequently observed in paediatric ALL. Hyperdiploidy is defined as greater than 50 chromosomes and is the most common genetic abnormality present in childhood ALL with an incidence of around 25-30% (Onciu 2009). Hyperdiploidy confers a favourable prognosis and current estimates of overall survival of children with hyperdiploid ALL is 90% (Paulsson & Johansson 2009). Furthermore, cases with trisomies 4, 10 and 17 have a particularly favourable outcome (Sutcliffe et al. 2005). Activating mutations affecting *FLT3* occur in around

20% of hyperdiploid ALL and efforts are underway to incorporate small molecule inhibitors of *FLT3* in standard therapy as these have been shown to be highly effective at killing *FLT3* positive, hyperdiploid ALL (Brown et al. 2005).

Hypodiploidy which is defined as fewer than 45 chromosomes, is much less frequent in childhood ALL, occurring in around 5% of cases of paediatric B-lineage ALL (Onciu 2009). Due to its low frequency, there have been few studies investigating the biological and clinical significance of this genetic subtype of ALL, however patients harbouring this genetic abnormality have a dismal outcome. A study conducted by Harrison and colleagues found that the 3yr EFS of 121 children with 42-45 chromosomes was around 65%, while those with 25-39 chromosomes (n=20) exhibited a 3yr EFS of only 30% (Harrison et al. 2004). As one might expect, deletions of B-cell developmental genes occurred in 100% percent of cases (Harrison et al. 2004).

1.3.2.6 Recurrent mutations in pre-B ALL

Genome-wide screens have provided insights into the molecular pathogenesis of ALL (Mullighan et al. 2007; Mullighan & Downing 2009). The most commonly disrupted loci in pre-B ALL are those encoding transcription factors involved in the regulation of B-cell development such as *PAX5*, *IKZF1* (*IKAROS*), *IKZF2* (*HELIOS*), *IKZF3* (*AIOLOS*), *LEF1*, *TCF3* and *BLNK*. Perturbations in lymphoid developmental genes occur in around 60% of cases of pre-B ALL and explain the developmental blockade that is a feature of this disease (Kuiper et al. 2007b; Mullighan et al. 2007). Surprisingly, DNA damage response pathways and tumour suppressor genes that are often mutated or lost in other tumour types are normally intact in pre-B ALL, with the prominent exception of the p16^{*INK4*}/*p14*^{*ARF*} locus which is altered in around 30% of paediatric pre-B ALLs (Kim et al. 2009).

Loss of function mutations affecting *IKAROS* (*IKZF1*) are associated with haematological malignancies including AML (Yagi et al. 2002) and pre-B ALL (Mullighan et al. 2007), and inactivating mutations or deletions affecting the *IKAROS*

locus are highly predictive of relapse in paediatric pre-B ALL (Kuiper et al. 2010; Mullighan et al. 2009). Ikaros proteins are nuclear factors required for normal haemopoiesis and lymphoid specification (Koipally et al. 1999; Cobb & Smale 2005). Mice homozygous for germline mutations in the murine Ikaros DNA-binding domain completely lack lymphoid and natural killer cells (Georgopoulos et al. 1994), while retaining normal development within the myeloid and erythroid compartment.

Ikaros proteins are sequence-specific transcription factors that have also been implicated in chromatin remodelling and can both repress and potentiate gene expression (Koipally et al. 1999; Rebollo & Schmitt 2003). Ikaros family proteins can form homodimers or heterodimers with other Ikaros family members (Rebollo & Schmitt 2003; Georgopoulos et al. 1994) and isoforms of the IKAROS protein generated by alternative splicing have been identified in both normal haemopoietic cells and leukaemic blasts (Tabayashi et al. 2007; Rebollo & Schmitt 2003; Iacobucci et al. 2008; Meleshko et al. 2008). These isoforms frequently lack part or all of the N-terminal zinc finger DNA-binding domain. However, the C-terminal dimerisation domain typically remains unaffected, thus these alternative isoforms can bind to canonical IKAROS (or other Ikaros family proteins) and act as dominant negative (DN) inhibitors (Sun et al. 1996).

1.3.2.7 Classical oncogenes associated with pre-B ALL

As well as activation of proto-oncogenes through cytogenetic rearrangements that are unique to haematological malignancies, there are also a number of “classical” oncogenes involved in leukaemogenesis, particularly in the case of *TEL-AML1* carriers who are thought to require a second oncogenic hit to progress to overt leukaemogenesis (Hong et al. 2008). One such example is the B-cell CLL/lymphoma gene (*BCL2*) (Vaux et al. 1988). *BCL2* activation is most frequently associated with t(14;18) translocations, however *BCL2* can also be overexpressed by activating point mutations or gene amplification. *BCL2* protein inhibits apoptosis by binding to and inhibiting the pro-apoptotic protein BCL2-associated X protein (BAX) (Zivny et al. 2010). While

expression of *BCL2* in ALL is common, it has not been associated with outcome (Gala et al. 1994).

One of the best studied oncogenes is the v-myc myelocytomatosis viral oncogene homolog, better known as *MYC*, which is overexpressed in a large number of solid tumours as well as some leukaemias (Albihn et al. 2010; Rubnitz & Crist 1997). T(8;14) rearrangements result in the translocation of the *MYC* gene to the immunoglobulin heavy chain (*IgH*) locus, resulting in an overexpression and an abundance of MYC protein driven by the *IgH* promoter.

MYC activity requires dimerisation with the basic helix-loop-helix (bHLH) protein MYC-associated factor X (MAX). The MYC-MAX dimer modulates gene transcription by binding to hexanucleotide sequences termed E boxes 5'CA[C/T]GTG-3' in the promoter of target genes (Luscher & Larsson 1999). MYC also interacts with proteins involved in chromatin remodelling such as transformation/transcription domain-associated protein (TRRAP), RuvB-like 1 (RUVBL1/TIP49) and RuvB-like 2 (RUVBL2/TIP48) (Amati et al. 2001; Wood et al. 2000; Park et al. 2001). A large number of genes are targets of MYC and many of these have been documented in the MYC Target Gene Database available online at: <http://www.myc-cancer-gene.org/site/mycTargetDB.asp>. To date this database contains 1697 individual genes that are predicted to be targets of MYC, and studies are ongoing to establish patterns of gene expression associated with oncogenic MYC activation (Albihn et al. 2010).

1.3.2.8 Tumour suppressor genes in pre-B ALL

Disruption of tumour suppressor genes (TSGs) are infrequent events in childhood ALL with the exception of those affecting the p16^{INK4}/p14^{ARF} locus which is altered in 25-30% of paediatric B-lineage ALLs by either mutation or hypermethylation (Krug et al. 2002; Kim et al. 2009). This locus encodes the cyclin-dependent kinase inhibitor 2A (*CDKN2A*) gene which can give rise to several different mRNA transcripts. Two *CDKN2A* variants encode protein isoforms of p16 that inhibit CDK4 kinase, while a

third *CDKN2A* transcript contains an alternative open reading frame (ARF) encoding the p14 protein which can inhibit MDM2. MDM2 binds to and induces degradation of the tumour suppressor p53, and p14 inhibits this interaction (Pomerantz et al. 1998). CDK4 and p53 both have prominent roles in regulating transition from G1 to S phase and are tightly regulated. Mutations involving the gene encoding p53 (*TP53*) are very common in solid malignancies, however the incidence of p53 mutations is extremely low in childhood cancers in general. Disruption of p53 is observed in around 5% of paediatric ALL cases, typically correlating with a poor outcome (Marks et al. 1997).

1.4 *CTGF* gene expression is dysregulated in pre-B ALL

A microarray study performed by Boag and colleagues compared the global gene expression profiles of 102 paediatric ALL samples with two control populations, bone marrow derived CD34^{pos} cells representing the HSC to pro-B cell compartment, and CD19^{pos}/IgM^{neg} cells representing the pro-B to small pre B-II compartment. It was expected that this approach would identify genes that may be relevant to leukaemogenesis (Boag et al. 2007). The results of this study demonstrated that one gene in particular, connective tissue growth factor (*CTGF*) displayed elevated expression, 19-fold higher vs CD34^{pos} and 37-fold higher vs CD19^{pos} IgM^{neg} cells in 75% of pre-B ALL samples. Expression was restricted to pre-B ALL and *CTGF* expression was not detected in T-ALL cells. Curiously, *CTGF* expression was also absent in the *TCF3-PBX1* cytogenetic group of pre-B ALL. Importantly, these findings were validated in an independent cohort of 132 paediatric ALL patients used in a study at St Jude Children's Research Hospital, Memphis (Ross et al. 2003). Deregulation of *CTGF* in B-lineage ALL has now been identified in four independent studies (Vorwerk et al. 2000; Sala-Torra et al. 2007; Boag et al. 2007; Kang et al. 2010). *CTGF* is not normally expressed by haemopoietic cells at any stage of development and has no known role in lymphocyte biology, thus the role of *CTGF* in pre-B ALL and the mechanism promoting its aberrant expression are unclear and remain to be elucidated.

1.5 Connective tissue growth factor

The name connective tissue growth factor (CTGF) was coined in 1991 by Bradham and colleagues to describe a novel protein secreted by a human umbilical endothelial cell line (HUVEC) that stimulated mitogenesis and chemotaxis of murine fibroblasts *in vitro* (Bradham et al. 1991). Subsequently, overexpression of CTGF has been observed in a number of fibrotic pathologies (Manetti et al. 2007; Lang et al. 2007; Leask 2004; Faull 2005). CTGF also has roles in numerous other biological and developmental processes including wound healing, angiogenesis, chondrogenesis, adhesion, apoptosis, chemotaxis, mitogenesis, extracellular matrix formation, endochondral ossification and tumorigenesis (Leask & Abraham 2006; Brigstock 2003; Dhar & Ray). CTGF appears to exert its effects through interacting with other protein factors through one of its four functional domains or by interacting with cell surface receptors and integrins. While some of these interactions have been characterised in various cell types, the precise role of CTGF in these biological contexts is largely unknown. Several recent reviews on CTGF function have highlighted the role of CTGF as a matricellular protein rather than a classical growth factor as CTGF is thought to facilitate interactions between cells and the surrounding extracellular matrix (Chen & Lau 2009; Brigstock 2010; Leask 2010). There is a growing body of evidence that CTGF acts to connect cells to the microenvironment by providing adhesive signals as well as modulating the availability of soluble growth factors, thus the tissue context within which CTGF is expressed is likely to play a key role in determining its biological function in any given microenvironment.

1.5.1 Origins

CTGF also known as CCN2, is a member of the CCN family of structurally related proteins, named after the three founding members Cysteine rich 61 (Cyr61 or CCN1), connective tissue growth factor (CTGF) and nephroblastoma overexpressed (NOV or CCN3) (Brigstock 2003). The remaining CCN members include the WNT1 inducible signalling pathway protein (WISP) members 1 to 3; WISP1 (CCN4), WISP2 (CCN5) and WISP3 (CCN6) (Leask & Abraham 2006). All six proteins share a similar modular

structure, as illustrated in Figure 1.2, containing up to four highly conserved domains. These domains contain motifs that resemble functional elements present in important regulatory proteins, including the insulin-like growth factor binding proteins (IGFBP), von Willebrand type-C (VWC), thrombospondin type-1 (TSP1) and a cysteine knot motif in the CT domain (Perbal 2001; Leask & Abraham 2006). This curious mosaic structure has resulted in CTGF (and other CCN proteins) being implicated both directly and indirectly in a vast array of biological processes and has made predicting the function of CTGF in any given biological context difficult (Chen & Lau 2009; Holbourn et al. 2008).

1.5.2 Protein structure

The CTGF protein contains 349 amino acid residues and has a molecular weight of 36-38 kDa depending on glycosylation of the full length protein (Bradham et al. 1991; Ball et al. 2003b). CTGF is a secreted protein and the N-terminus contains a 37 amino acid signal peptide (SP) directing the protein through the Golgi apparatus to the cell membrane for secretion (Abreu et al. 2002; Chen et al. 2001). CTGF is thought to have a short half-life and is susceptible to proteolysis by several enzymes, including elastase, plasmin and the matrix metalloproteases (MMPs) 1, 3, 7 and 13 (Hashimoto et al. 2002). Several smaller molecular weight fragments of CTGF have been detected in biological fluids and these are thought to potentiate specific biological functions (Yang et al. 1998; Steffen et al. 1998; Ball et al. 1998). These smaller isoforms, which arise from proteolysis upon secretion, range in size from 10-20 kDa and are primarily comprised of either module 4 alone or modules 3 and 4 combined (Brigstock et al. 1997). The two enzymes elastase and plasmin can cleave CTGF between any of the four modules (Hashimoto et al. 2002), while MMPs 1, 3, 7 and 13 cleave the CTGF protein in the proteinase-sensitive hinge region between modules 2 and 3 (Gressner & Gressner 2008).

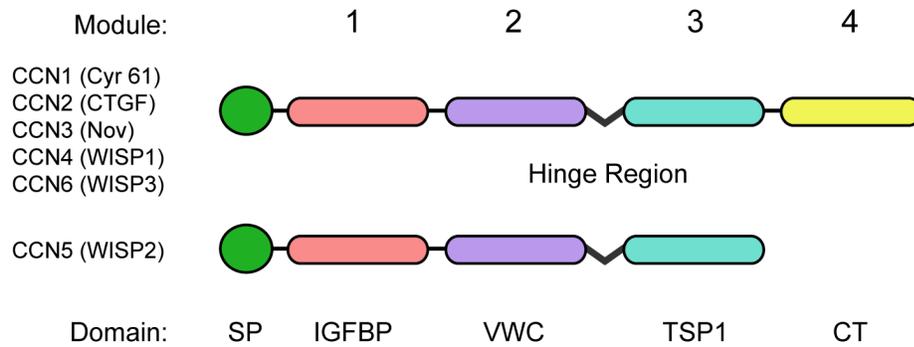


Figure 1.2 Modular structure of CCN proteins

All CCN members share a similar modular structure including an N-terminal signal peptide (SP), insulin-like growth factor binding protein domain (IGFBP), Von Willebrand type-C (VWC) domain, and a thrombospondin type-1 domain (TSP1). CCNs1-4 and CCN6 contain a cysteine-rich C-terminus (CT) which is not present in CCN5. All CCN family members contain a protease-sensitive hinge region between module 2 and 3. Figure adapted from Leask, A. and Abraham, D. (2006).

1.5.3 CTGF signalling

CTGF is pleiotropic and has been implicated in numerous biological processes, largely due to the diversity of proteins shown to interact with CTGF and other CCN family members. Our understanding of the various roles CTGF may play in any given microenvironment or biological process is limited at present and important questions that require addressing include which cellular receptors can be activated by CTGF, which soluble factors can CTGF bind to, and can these interactions modulate the function of target proteins? The following sections outline some of the key protein factors and receptors that are known to bind CTGF.

1.5.3.1 Insulin signalling

Module 1 of CTGF contains a sequence motif homologous to the IGFBP superfamily and this domain has been shown to bind human recombinant insulin-like growth factor (IGF) 1 and 2 with low affinity (Kim et al. 1997). IGF family proteins have been

implicated in haemopoietic development and importantly in leukaemia (Zumkeller 2002; Elmlinger et al. 1996). Furthermore, the first observation that *CTGF* was expressed specifically in pre-B ALL came from a study investigating the expression of insulin binding protein genes in leukaemia (Vorwerk et al. 2000). However, there is little functional evidence so far that CTGF plays a significant role in IGF signal transduction and the biological implications of the structural relationship between CTGF and IGFbps remains unclear and warrants further investigation.

1.5.3.2 TGFB family signalling

CTGF module 2 contains a cysteine rich VWC motif and this domain can bind members of the transforming growth factor beta superfamily of proteins including; bone morphogenic proteins (BMPs) 4 and 7 and transforming growth factor beta 1 (TGFB1) (Abreu et al. 2002; Nguyen et al. 2008). Binding modulates the availability of these factors to attach to their cognate receptors and activate intracellular signalling. Interaction between CTGF and either BMP4 or BMP7 impairs the ability of these proteins to bind to the type II TGF receptor (TGFB2). Conversely, TGFB1 bound to CTGF displays enhanced TGFB2 binding (Abreu et al. 2002). Interestingly, CTGF can be potently induced by TGFB1 signalling and this is the best studied inducer of *CTGF* gene expression (Trojanowska 2009; Arnott et al. 2008). Furthermore, around a third of TGFB1-responsive genes are no longer inducible in *CTGF* null embryonic fibroblasts, demonstrating the critical role CTGF plays in signalling downstream of TGFB1 (Shi-wen et al. 2006). These data indicate that the CTGF/TGFB1 axis may demonstrate a positive feedback loop.

1.5.3.4 Vascular endothelial growth factor signalling

CTGF has been suggested to play a role in regulating the angiogenic effects of vascular endothelial growth factor (VEGF). Inoki and colleagues identified CTGF in a screen for VEGF-binding partners using a yeast two-hybrid system (Inoki et al. 2002). The authors further demonstrated that CTGF has the ability to bind to VEGF, using either module 3

(TSP1 domain) or module 4 (CT domain) independently. Full length CTGF sequesters VEGF in a complex preventing it from binding to cell surface receptors, and experiments have shown that cleavage of this complex by a number of MMPs in the CTGF hinge region can result in the release of previously bound VEGF (Hashimoto et al. 2002; Ito et al. 2007). This mechanism is suggested to fine-tune the bioavailability of VEGF. Furthermore, MMPs are commonly expressed and activated during wound healing and tissue remodelling (Murphy & Nagase 2008), two environments where angiogenesis is required. An overall increase in cell surface or extracellular expression of MMP family proteins is likely to impair the ability of CTGF to bind VEGF and induce the release of any VEGF that is sequestered in either stromal or cell surface-bound complexes. Curiously, reminiscent of TGF β 1, induction of *CTGF* gene expression by VEGF has also been reported (He et al. 2003; Suzuma et al. 2000; Kuiper et al. 2007a), however these findings were obtained in studies of bovine retinopathies and may not represent the role for CTGF in human biology, and further investigation is required.

1.5.4 CTGF cell surface interactions

As well as binding important regulatory proteins, CTGF has been shown to bind several membrane-bound receptors that can facilitate intracellular signalling *via* tyrosine phosphorylation. Because the response to extracellular CTGF appears to vary depending on the tissue context, an important step in elucidating the potential role of CTGF in any given microenvironment is to determine what receptors are available to bind CTGF and what cells are expressing such receptors. Some of the known CTGF receptor interactions that have been validated are outlined below.

1.5.4.1 LRP1 receptor

A study performed by Segarini and colleagues demonstrated the ability of CTGF to bind the low density lipoprotein receptor-related protein (LRP1), also known as α 2-macroglobulin receptor (α ₂MR) through interactions with the TSP1 domain in module 3

of the CTGF protein (Segarini et al. 2001). In this study, the authors used LRP1 purified from lysates prepared from the bone marrow stromal cell line BSM2, suggesting that LRP1 has the potential to act as a CTGF receptor in the bone marrow microenvironment. LRP1-CTGF interactions have also been confirmed in rat hepatic stellate cells and human renal fibroblasts (Gao & Brigstock 2003; Yang et al. 2004). Yang and colleagues confirmed phosphorylation of the intracellular domain of LRP1 in response to CTGF binding, resulting in activation of downstream mitogen-activated protein kinase (MAPK) signalling (Yang et al. 2004). Blockade of LRP1 with receptor associated protein (RAP) or inhibition of MAPK activation with the MAPK/ERK kinase 1 (MEK1) inhibitor PD98059 reduced myofibroblast differentiation in response to exogenous CTGF (Yang et al. 2004). CTGF has also been shown to bind the Wnt-co receptor LRP6 on *Xenopus* embryos, inhibiting Wnt signaling. However, this has not been investigated in mammalian cells (Mercurio et al. 2004).

1.5.4.2 NTRK1/NGFR receptor complex

CTGF is thought to play a role in diabetic retinopathy and a study by Wahab et al demonstrated that CTGF activates signalling pathways, including ERK 1/2, JNK, ERK5, phosphatidylinositol 3-K, CaM-KII, PKC α , and PKC δ in human mesangial cells (Wahab et al. 2005). Furthermore they demonstrated that this response was downstream from CTGF binding to the type 1 neurotrophic tyrosine kinase receptor (NTRK1/TrKA) and the co-receptor nerve growth factor receptor (NGFR/p75^{NTR}). The cytoplasmic domain of NTRK1 contains tyrosine residues that can be phosphorylated and interact directly with intracellular adapter proteins, resulting in a vast array of biological responses, including proliferation, cytoskeletal remodelling and membrane trafficking, and differentiation (Huang & Reichardt 2003).

1.5.4.3 CTGF-associated adhesion

Module 4 of CTGF (CT domain) binds heparin sulphate proteoglycans (HSPGs). These proteins are expressed on the surface of some cell types and HSPG-CTGF interactions

have been demonstrated in fibroblasts (Chen et al. 2000), hepatic stellate cells (Gao & Brigstock 2004), and some monocytes (Schober et al. 2002). In each case, specific integrin complexes were required as co-receptors and these differed between cell types, including $\alpha\text{m}\beta\text{2}$ integrins in monocytes, $\alpha\text{6}\beta\text{1}$ in fibroblasts and $\alpha\text{5}\beta\text{3}$ in hepatic stellate cells. Furthermore binding of CTGF to $\alpha\text{5}\beta\text{3}$ integrins through the CT domain facilitates migration and adhesion of chondrosarcoma cells independent of HSPG binding (Tan et al. 2009). Other integrins known to interact with CTGF include $\alpha\text{5}\beta\text{1}$ on pancreatic stellate cells and chondrocytes (Gao & Brigstock 2006; Hoshijima et al. 2006; Nishida et al. 2007), and $\alpha\text{IIb}\beta\text{3}$ on human platelets (Jedsadayamata et al. 1999). Hoshijima and colleagues have demonstrated that fibronectin (FN1) also binds the CT domain, promoting adhesion of chondrocytes. This process was dependant on co-binding $\alpha\text{5}\beta\text{1}$ integrins (Hoshijima et al. 2006).

HSPGs are common components of stroma and this has led to the hypothesis that the microenvironment may act as a sink for bioactive CTGF. Perlecan is the major extracellular matrix proteoglycan in the bone marrow (Schofield et al. 1999) and has been shown to sequester CTGF as well as a number of other important protein factors, including fibroblast growth factors (FGFs), platelet derived growth factor (PDGF), VEGF and many more (Melrose et al. 2008). It is not known whether CTGF bound to stroma can interact with other binding partners such as TGFB1, BMPs or VEGF and careful functional characterisation of HSPG-bound CTGF is required before such fundamental biological questions can be answered.

1.5.5 The *CTGF* locus

The human *CTGF* gene maps to the long arm of chromosome 6, at cytoband q23.1 (Martinerie et al. 1992). *CTGF* is an immediate early gene and can be expressed rapidly in the absence of endogenous protein synthesis, highlighting its importance in mammalian homeostasis (Igarashi et al. 1993). The *CTGF* gene is relatively small at approximately 3.5 kb in length and encodes a single mature mRNA transcript of 2.4 kb in length, containing five exons. There have been no conclusive reports of *CTGF* splicing variants or alternative transcripts, however a recent study investigating the

exome of thymic tumours identified *CTGF* mRNA transcripts with reduced inclusion of exon 2 (Soreq et al. 2008). Follow-up studies need to be performed to accurately characterize *CTGF* mRNA transcripts in these samples.

1.5.5.1 The *CTGF* promoter

CTGF expression can be modulated by a number of extracellular stimuli, including high glucose (Wang et al. 2008), hypoxia (Higgins et al. 2004; Hong et al. 2006), mechanical stress (Nishida et al. 2008) and soluble factors such as transforming growth factor beta (TGFB1) (Xie et al. 2005; Arnott et al. 2008), endothelin-1 (ET-1) (Recchia et al. 2009), serum response factor (SRF) (Muehlich et al. 2007) and curiously the matrix metalloprotease MMP3 (Eguchi et al. 2008). Other factors have also been implicated in the induction of *CTGF* expression, however these observations are largely cell type specific and in some cases indirect, requiring further investigation. The *CTGF* promoter contains several well characterised regulatory elements, outlined in Figure 1.3. These include binding sites for Sp1 and Ap-1/c-Jun (Holmes et al. 2003; Yu et al. 2009), tandem Ets-1 binding motifs (Van Beek et al. 2006), as well as a Smad-binding element (SBE) (Holmes et al. 2001), and the basal control element (BCE-1), originally named the TGFB1 response element (TBRE) (Grotendorst et al. 1996).

TGFB1 has long been recognised as one of the most potent inducers of *CTGF* expression in fibroblasts, mesangial cells and vascular smooth muscle cells (Grotendorst 1997; Chen et al. 2002; Fu et al. 2001). The BCE-1 was originally described by Grotendorst and colleagues as a promoter region critical for the induction of *CTGF* expression by TGFB1. This promoter element is activated downstream of MEK/ERK signalling and mutation of this region can abolish TGFB1-induced *CTGF* expression (Chen et al. 2002; Holmes et al. 2001; Pickles & Leask 2007). The SBE is also critical for TGFB1-induced *CTGF* expression. Signalling through the type II TGFB receptor (TGFB2) results in activation of smad2 and smad3 nuclear factors. These factors in turn dimerize with smad4 to activate gene transcription through binding at the SBE. Smad3 and smad4, but not smad2 have been shown to be critical for TGFB1-induced

CTGF expression in fibroblasts, osteoblasts, and proximal-tubule epithelial cells (Holmes et al. 2001; Arnott et al. 2008; Phanish et al. 2006).

The SBE can also act in concert with the tandem Ets-1 sites and this synergy is required for the induction of *CTGF* by TGF β 1 in fibroblasts (Van Beek et al. 2006). These tandem Ets-1 binding sites have also been shown to be a target of the oncogenic transcriptional co-activators YAP and TEAD, and *CTGF* expression is potently activated by YAP in NIH-3T3 and MCF10A breast cancer cells (Zhao et al. 2008). An element termed the transcriptional enhancer dominant in chondrocytes (TRENDIC) can be activated by the matrix metalloprotease MMP3 (Eguchi et al. 2008). Transcriptional regulation is not a function commonly attributed to matrix metalloproteinases, however Eguchi and colleagues demonstrated nuclear localisation of MMP3 in chondrocytes and confirmed trans-activation of *CTGF* gene expression by nuclear MMP3 bound to the *CTGF* promoter (Eguchi et al. 2008). *CTGF* expression can also be modulated by hypoxia through two HIF response elements (HRE) at -3745 and -1558 (Higgins et al. 2004), however hypoxia has been shown to induce (Samarin et al. 2010) and inhibit (Kroening et al. 2009) *CTGF* expression depending on cell type and activation of different signalling pathways downstream from hypoxic stimuli, thus the role of hypoxia in regulating *CTGF* expression is controversial. Because *CTGF* is not normally expressed in lymphoid cells, it is unclear what transcriptional mechanisms are likely to be responsible for its aberrant expression in pre-B ALL.

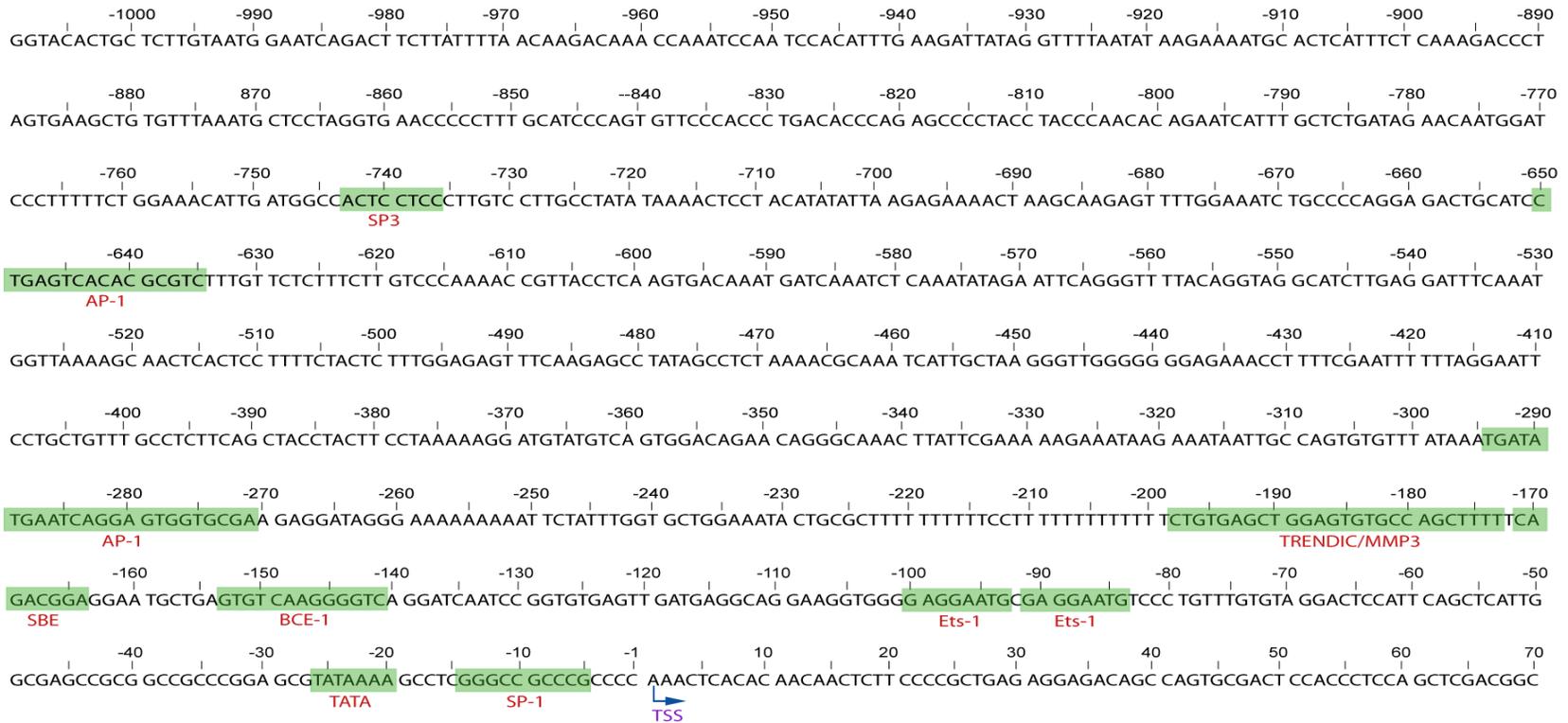


Figure 1.3 Transcription factor binding sites at the *CTGF* promoter

The *CTGF* promoter region from -1009 to +70 is shown. Functional transcription factor binding sites (green boxes) include two AP-1 sites (Yu et al. 2009), the transcriptional enhancer dominant in chondrocytes (TRENDIC) which is bound by MMP3 (Eguchi et al. 2007; Eguchi et al. 2008), an SP3 binding site (Fonseca et al. 2007), a smad binding element (SBE) (Arnott et al. 2008), basal control element (BCE-1) (Grotendorst 1997), tandem Ets-1 binding sites (Van Beek et al. 2006), a TATA box and an SP-1 binding site (Holmes et al. 2003). TSS: transcription start site (+1).

1.5.5.2 Epigenetic regulation of *CTGF* expression

Epigenetic features are now widely acknowledged to be important in tumorigenesis (Bonifer & Bowen 2010). Post-translational modifications of histone tails can alter chromatin structure, impacting on the accessibility of DNA elements for transcriptional initiation and elongation (Chi et al. 2010). Methylation of DNA at CpG dinucleotides can directly alter the ability of trans-factors to bind to DNA and can also recruit protein complexes that can remodel chromatin and promote gene inactivation, thus an increase in DNA methylation at gene promoters is generally associated with inactivation of gene transcription (Watanabe & Maekawa 2010). Cancer cells frequently exhibit global hypomethylation of DNA with an accompanying focal, hypermethylation at tumour suppressor loci (Kondo & Issa 2010). The role of epigenetic factors in the regulation of *CTGF* gene expression is still not clear, however several studies have identified cancer associated changes in DNA methylation and histone acetylation affecting the *CTGF* locus (Komorowsky et al. 2009; Chiba et al. 2004; Kikuchi et al. 2007). Chiba and colleagues identified an association between *CTGF* gene expression and histone acetylation, as well as changes in DNA methylation in some but not all hepatoma cell lines (Chiba et al. 2004; Chiba et al. 2005). An association between histone acetylation and *CTGF* expression has also been observed in renal endothelial cells (Komorowsky et al. 2009). DNA methylation of this region is inversely correlated with *CTGF* gene expression in ovarian cancer specimens and cell lines (Kikuchi et al. 2007). These findings suggest that aberrant epigenetic regulation may play a role in inducing *CTGF* expression in some cancers.

Recent evidence points to a crucial role for DNA methylation in haematological malignancies more generally. Milani and colleagues assessed the methylation status of 1320 CpG sites in the regulatory region of 416 genes and showed that the methylation status of 300 of these CpG sites facilitated the accurate prediction of lineage and cytogenetic subtype in a large cohort of paediatric ALL specimens (Milani et al. 2010). Furthermore, a recent analysis of the global DNA methylation status of 367 haematological neoplasms by Martin-Subero and colleagues identified increased

methylation of the *CTGF* locus in T-ALL and conversely a reduction in methylation of this locus in B-lineage ALLs (Martin-Subero et al. 2009). Given the notable absence of *CTGF* expression in T-ALL, this suggests that DNA methylation may be involved in regulating the *CTGF* locus in ALL (Martin-Subero et al. 2009). This study did however have significant limitations as only 1505 individual CpG sites were analysed, spread across 807 genes, and only 2 CpG residues were interrogated at the *CTGF* locus. Thus, a thorough investigation into epigenetic regulation of *CTGF* expression in ALL is warranted.

1.5.5.3 The *CTGF* 3' UTR

The 3' untranslated region (3' UTR) of canonical *CTGF* mRNA contains important regulatory elements and several miRNAs have been functionally demonstrated to regulate *CTGF* post-transcriptionally by inducing degradation of *CTGF* mRNA. These include miR18a, miR30, miR130 and the miR-17-92 cluster (Ohgawara et al. 2009; Duisters et al. 2009; Ernst et al. 2010). Another sequence element termed cis-acting element of structure-anchored repression (CAESAR) by Kubota and colleagues, has been shown to bind an as yet unidentified nuclear factor in chondrocytes, resulting in attenuation of CTGF protein synthesis (Kubota et al. 2000). Disruption of this sequence influences the stability of the *CTGF* mRNA, particularly in response to hypoxic conditions (Kubota et al. 2000; Kondo et al. 2006; Kubota et al. 2005). This element has the potential to act as a powerful inhibitor of CTGF protein expression and may serve as a negative regulator in other tissue types. While it is possible that modulation of *CTGF* mRNA stability conferred by the CAESAR element is mediated by miRNA, no functional miRNA binding domains have been identified in this 84bp sequence element.

1.5.6 *CTGF* expression and cancer

Aberrant expression of *CTGF* has been implicated in the biology of up to 21 different cancer types to date (Chu et al. 2008). However, the precise role that CTGF plays in tumour biology is presently unclear. This is highlighted by the observations that CTGF

can both promote and inhibit cancer progression, depending on the type and location of the tumour. *CTGF* expression has been associated with an aggressive phenotype in breast cancer (Kang et al. 2003), adenocarcinoma (Koliopoulos et al. 2002), pancreatic cancer (Wenger et al. 1999), malignant melanoma (Kubo et al. 1998) and glioblastoma (Yin et al. 2010). Conversely, *CTGF* expression has been associated with reduced metastasis and proliferation in ovarian cancer (Barbolina et al. 2009), chondrosarcoma (Shakunaga et al. 2000), lung cancer (Chien et al. 2006) and squamous cell carcinoma (Moritani et al. 2003). This duality likely reflects the importance of biological context and microenvironment on the action of *CTGF*.

CTGF expression has been shown to have prognostic relevance in ALL. A study investigating patient outcomes in adult ALL identified an association between high *CTGF* mRNA expression and a poor outcome (Sala-Torra et al. 2007). Patients were stratified based on the level of *CTGF* mRNA expression in their leukaemic blasts into three groups, low (n=26), intermediate (n=26) or high (n=26). The five year event free survival rates for these groups were 58%, 42% and 12% respectively (Sala-Torra et al. 2007). More recently, a COG study investigating gene expression classifiers for improved prognostic stratification of high-risk paediatric pre-B ALL patients identified an association between high *CTGF* expression and low relapse free survival (RFS) (Kang et al. 2010). Clearly, high *CTGF* expression is associated with a poor outcome in ALL and studies to establish the biological role of *CTGF* in this disease are warranted.

1.5.7 *CTGF* gene expression in the bone marrow

The bone marrow space is exposed to *CTGF* during normal development and maintenance, and *CTGF* is required for differentiation and proliferation of chondrocytes and osteoblasts (Takigawa et al. 2003; Nishida et al. 2000). The process of endochondral ossification whereby cartilaginous framework is replaced by bone, is distorted in the absence of *CTGF* protein (Song et al. 2007; Takigawa et al. 2003), and *CTGF* knock out mice die shortly after birth due to respiratory failure, as a result of significant skeletal

defects (Ivkovic et al. 2003). Other potential sources of CTGF protein in the bone marrow include fibroblasts, bone marrow stromal cells and endothelial cells.

Evidence supporting a role for *CTGF* expression in normal haemopoiesis has not been observed, however a 2007 study provided tantalizing data demonstrating that *CTGF* gene expression is induced in CD34^{pos} haemopoietic progenitor cells (HPCs) by adhesion to bone marrow stromal cells *in vitro* (Wagner et al. 2007). Adherent HPCs displayed 60-fold higher *CTGF* mRNA expression compared to non-adherent HPCs. Additional investigation is required to delineate the specific cell types expressing *CTGF* in this model system, as the CD34 marker can also be expressed by other cell types present in the bone marrow such as fibrocytes and vascular endothelial cells (Barth & Westhoff 2007; Nielsen & McNagny 2009). Nevertheless, this study points to a role for CTGF in adhesion of CD34^{pos} HPCs in the bone marrow and is the first such evidence of a role for CTGF in haemopoietic development.

1.5.8 CTGF, ALL and bone marrow fibrosis

High CTGF levels feature prominently in fibrotic pathologies and much research is directed toward understanding the role of CTGF in fibrotic disease (Wang et al. 2011; Brigstock 2010). In 1964 a study identified a concordance between bone marrow fibrosis and B-lineage ALL (Kundel et al. 1964), and this finding was further reinforced in 1989, with the reported observation that bone marrow fibrosis was present in 46 of the 56 (82%) paediatric ALL cases of B-lineage, but only 1 out of 7 patients with T-lineage ALL (Wallis & Reid 1989). Confirmation as to the importance of these early studies has been forthcoming with a report from Sweden by Noren-Nystrom and colleagues confirming increased bone marrow reticulin fiber density (RFD), a marker of fibrosis, in pre-B ALL versus T-ALL specimens (Noren-Nystrom et al. 2008). Furthermore, the clinical relevance of this increased bone marrow fibrosis was highlighted by the finding that the level of RFD at diagnosis correlated with the level of MRD after induction chemotherapy, suggesting that fibrotic bone marrow provides protection to leukaemic cells against chemotherapy (Noren-Nystrom et al. 2008). While this apparent fibrotic

pathology cannot be attributed retrospectively to the action of CTGF, the incidence of fibrosis and the predilection toward B-lineage ALL is consistent with the observed incidence of aberrant *CTGF* expression in pre-B ALL (~75%) (Boag et al. 2007; Vorwerk et al. 2000).

Solid tumours are now known to be comprised of not only neoplastic cells, but also stromal cells, including endothelial cells, inflammatory cells and fibroblasts (Chouaib et al. 2011; Allen & Louise Jones 2011). Cancer associated fibroblasts (CAFs) can enhance tumour growth and metastasis through release of classical growth factors, chemokines and enhanced synthesis of extracellular matrix (Kalluri & Zeisberg 2006). Elevated CTGF protein expression has been detected in the stroma of a number of tumours, including pancreatic, breast and oesophageal cancers (Wenger et al. 1999; Frazier & Grotendorst 1997; Koliopanos et al. 2002), suggesting an important role for CTGF in modulating the tumour microenvironment. In a xenograph model of human prostate cancer, human prostate carcinoma cells (LNCaP) were combined with murine fibroblasts carrying an inducible *CTGF* transgene. Expression of *CTGF* in murine fibroblasts significantly enhanced LNCaP engraftment and tumour growth, suggesting CTGF plays an important role in the tumour-promoting functions of CAFs (Yang et al. 2005).

While the role of CAFs in the growth and maintenance of solid tumours is now well established, the role of bone marrow stromal cells in promoting leukaemogenesis remains ill-defined. A number of stromal cell types exist in the bone marrow and these cells together with osteoblasts play important roles in regulating specialised HSC niches (Wilson & Trumpp 2006; Wilson et al. 2009). Furthermore, recent studies by Lee and colleagues have shown that CTGF can induce terminal differentiation of primary human bone marrow stromal cells into fibroblasts (Lee et al. 2010), thus ectopic expression of CTGF in the bone marrow microenvironment is likely to significantly alter the bone marrow stroma and may provide ALL cells with a growth advantage or refuge from chemotherapy (Lane et al. 2009).

1.6 Overall objectives of this study

Regulation of *CTGF* expression is complex and mechanisms vary significantly between cell types (Eguchi et al. 2007; Chaqour & Goppelt-Struebe 2006; Shi-Wen et al. 2008). While transcriptional networks have been elucidated in some cell types, the mechanisms by which *CTGF* expression is regulated in haemopoietic cells or leukaemia has not yet been investigated, as *CTGF* is not normally expressed in cells of haemopoietic origin. Transcription factors, epigenetic factors and miRNAs may all play a role in regulating *CTGF* expression and any number of these may be responsible for the observed dysregulation of the *CTGF* locus in pre-B ALL.

The specific aims of this thesis were as follows:

- i. To examine whether aberrant *CTGF* expression is a result of abnormal transcription factor expression or activation.
- ii. To examine clinical features associated with *CTGF* gene expression in pre-B ALL, and furthermore to investigate the association between *CTGF* expression and overall patient outcomes.
- iii. To determine if aberrant *CTGF* expression is a result of genomic lesions such as structural rearrangements or gene mutations.
- iv. To determine whether aberrant *CTGF* expression has an epigenetic basis.

Chapter 2

Materials and Methods

Chapter 2

Materials and Methods

2.1 Specimens and cell lines

2.1.1 Patient specimens

Bone marrow specimens were obtained from children diagnosed with ALL at Princess Margaret Hospital for Children, Perth, WA, Australia, between 1984 and 2005. Patients were treated according to Children's Cancer Group (CCG) protocols (Gaynon et al. 2000), and specimens were obtained either at the time of initial diagnosis, relapse or both. Ethics approval for this study was obtained from the Institutional Review Board, and informed consent for the use of tissues for research purposes was obtained for all individuals. White blood cells (WBC) were separated from whole BM specimens at the time of diagnosis using a Lymphoprep gradient (Nycomed) according to the manufacturer's instructions, then cryopreserved.

2.1.2 Donor mononuclear cells

Peripheral blood mononuclear cells (PBMCs) were isolated from healthy adult donors after provision of informed consent. Briefly, 8 ml of donor blood was collected into heparin vacutainers (BD Bioscience), inverted several times and left at room temperature for 15 min. Blood was separated by centrifugation at 1500 x g for 15 mins at 4°C. After centrifugation, the buffy coat was collected and adjusted to 10 ml with ice-cold PBS, washed twice with a further 10 ml of ice-cold PBS, then resuspended in one pellet volume of ice-cold PBS. PMBCs were processed immediately for nucleic acid extraction (described below), or stored at -80°C for no more than 2 weeks.

2.1.3 Mammalian tissue culture

The T-ALL cell line Jurkat was obtained from the American Type Tissue Collection (ATCC). All remaining ALL cell lines used in this study were generated at the Telethon

Institute for Child Health Research in the laboratory of Prof Ursula Kees according to previously published methods (Kees et al. 1987). Culture medium used for all cell culture was RPMI supplemented with 1 mM sodium pyruvate (MP Biomedicals, LLC), non-essential amino acids (MP Biomedicals, LLC), 2 nM 2-mercapto-ethanol, and 2 mM L-glutamine. Culture medium used for PER-145, PER-278, PER-371 and PER-377 was supplemented with 20% fetal calf serum (FCS) (Invitrogen, Life Technologies), while culture medium used for PER-485, PER-490, PER-495 and JURKAT was supplemented with 10% FCS. Cells were maintained in 5% CO₂ at 37°C.

2.1.3.1 Culture with 5-aza-2-deoxycytidine and Trichostatin A

Cells were seeded in fresh culture medium at 10⁶ cells/well in 24 well culture plates (Nunc, Thermo Fisher Scientific). In culture, 5-aza-2-deoxycytidine (Aza) was added to medium to a final concentration of 5 µM. Cells were treated for a total of 72 h with medium containing Aza replaced daily. Cells treated with Trichostatin A (TSA) were grown for 72 h, but treated for only the last 12 h in culture with TSA at a concentration of 330 nM. For cells treated with both agents, medium containing 5 µM Aza and 330 nM TSA was added to cells 60 h after the first Aza treatment.

2.2 Bioinformatics and statistics

2.2.1 Microarray data analysis

Global gene expression data from the PMH cohort was obtained by Boag and colleagues using the human Affymetrix U133A platform as described previously (Boag et al. 2007; Boag et al. 2006). Global gene expression data for the Ross cohort was obtained using the Affymetrix U133A platform according to the methods reported by Ross and colleagues (Ross et al. 2003). The Kang cohort was analysed using the Affymetrix plus 2.0 platform and global expression data was normalised according to methods outline in the report by Kang and colleagues (Kang et al. 2010). Normalised gene expression data for the Kang cohort was obtained from the European Bioinformatics Institute's Array

Express website <http://www.ebi.ac.uk/arrayexpress/> under the accession number E-GEOD-11877. Gene expression data for the Ross cohort was provided as supplemental data in the original publication (Ross et al. 2003).

2.2.1.1 Establishing gene expression cut-offs

Micro-array gene expression cut-offs were established based upon comparisons between global gene expression data from the PMH cohort and qRT-PCR validation experiments (Boag et al. 2007). Robust multi-array analysis (RMA) normalised probeset signals of < 50 were classed as negative for the mRNA targeted by that probeset. For *CTGF* mRNA expression levels, specimens were classed as high if the *CTGF* probeset intensity was > 500, med if between 500 and 50, and low/neg if < 50.

2.2.2 Promoter analysis

The MEME software suite (Bailey & Elkan 1994) was used to compare gene promoter regions for shared sequence motifs. Common motifs identified by analysis with MEME were screened for similarities with annotated transcription factor binding matrices contained in the JASPER database (Portales-Casamar et al. 2010) using the TOMTOM web interface (Gupta et al. 2007). Promoter regions were also manually scanned for transcription factor binding sites using the TFSEARCH database of transcription factor binding sites (Akiyama n.d.), and the PROMO transcription factor binding prediction software (Farre et al. 2003).

2.2.3 Statistical analyses

2.2.3.1 Spearman correlation analyses

Spearman's correlations were performed by Dr Martin Firth at the Telethon Institute for Child Health Research, Subiaco, Western Australia using the R software environment.

For each probeset contained within the three microarray datasets (PMH, Kang and Ross), its correlation with the *CTGF* probeset (209101_at) was calculated. This analysis was performed independently for all three microarray cohorts and probesets were ranked based upon correlation values. The top 100 *CTGF*-associated probesets were compared amongst cohorts.

2.2.3.2 Pearson correlation analyses

CTGF gene expression was measured in a selection of primary specimens from the PMH cohort by real-time PCR (section 2.3.4.1). These expression values were compared to those generated by microarray analysis of matched specimens in the PMH cohort described previously (Boag et al. 2007). Concordance was examined between these methods by carrying out a Pearson correlation analysis. This was performed using the “CORREL” function in the Excel spreadsheet application (Microsoft).

2.2.3.3 Graphing and statistical analysis of gene expression data

Graphing of all gene expression, patient survival and densitometry data was performed using Prism 4.0 (GraphPad Software). Statistical comparison of means was performed by t-tests or ANOVA in Prism 4.0.

2.3 Molecular Techniques

Unless otherwise stated, all biochemical and molecular reagents were obtained from Sigma Aldrich. A full list of suppliers is provided as an appendix at the end of this chapter in Table A2.1. A list of primers and oligonucleotides is presented as an appendix of this chapter in Table A2.2.

2.3.1 Isolation of nucleic acids

Prior to extraction of nucleic acids from bone marrow specimens or PMBCs, cryopreserved specimens were thawed rapidly at 37°C, transferred to ice, then processed according to extraction protocols below. Quantitation of nucleic acids was performed using an ND1000 spectrophotometer (ThermoFisher Scientific).

2.3.1.1 Genomic DNA extraction

After thawing of cryopreserved primary specimens or collection of cells from culture, 5×10^6 cells were washed twice in ice-cold PBS and then genomic DNA was extracted using the DNeasy Blood and Tissue kit (QIAGEN) as per the manufacturer's instructions. DNA was eluted in TE buffer and stored at 4°C until use.

2.3.1.2 Total RNA extraction

Total RNA was isolated from cells using a modified TRIZOL method (Hoffmann et al. 2005). After two washes with ice-cold PBS, cells were resuspended in one pellet volume of PBS and combined with 0.5 ml TRIZOL reagent (Invitrogen, Life Technologies) per 10^6 cells. Samples were vortexed thoroughly and left at room temperature for 5 min. After the addition of 0.1 ml chloroform, samples were vortexed thoroughly and centrifuged at 15,000 x g for 15 min at 4°C. The aqueous phase was then transferred to fresh nuclease-free, sterile 1.5 ml tubes and combined with 0.54 vol of absolute EtOH and vortexed. The resulting solution was purified using the RNeasy Mini kit (QIAGEN) following the manufacturer's instructions, from step 5 of the protocol: "Purification of total RNA from animal cells" in the RNeasy mini kit handbook. On-column DNase treatment was included during purification. Samples were eluted in 50 µl of nuclease free H₂O, quantitated and stored immediately at -80°C.

2.3.1.3 Cytoplasmic RNA extraction

Cytoplasmic RNA for northern blotting was isolated from ALL cell lines by the following method. Cells were collected by centrifugation at 3,000 x g then washed twice

with 10 ml of ice-cold PBS and resuspended in 200 μ l of lysis buffer [10mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM MgCl₂, 0.5% nonident P-40] per 5 x 10⁶ cells and incubated on ice for 10 min. Nuclei were collected by centrifugation at 12,800 x g for 10 min at 4°C. Supernatant containing cytoplasmic RNA was collected into fresh sterile, nuclease-free 1.5 ml tubes and combined with 25 μ l of 10% sodium dodecyl sulphate (SDS) on ice. RNA was extracted twice with 100 μ l of buffered phenol solution. After addition of the phenol solution, samples were vortexed thoroughly and centrifuged at 15,000 x g for 3 min at room temperature, and supernatants were transferred to fresh 1.5 ml tubes. Following the second phenol extraction, supernatants were combined with 0.1 vol of 2 M sodium acetate (pH 5.0) and 2.5 vol of absolute EtOH and incubated at -20°C overnight. RNA was pelleted by centrifugation at 15,000 x g for 15 min at 4°C. Pellets were washed with 70% EtOH, air dried briefly, and then resuspended in 50 μ l of nuclease-free H₂O and stored immediately at -80°C.

2.3.2 Polymerase chain reaction

Polymerase chain reaction (PCR) assays were performed on either a PTC 200 (MJ Research), or a C1000 Thermal Cycler (Bio-Rad). All PCR reactions were assembled on ice in sterile 0.2 ml thin-walled tubes (Bio-Rad). Primer and oligonucleotide sequences are listed in Table A2.

2.3.2.1 Standard polymerase chain reaction

Standard PCRs were performed using GoTaq Flexi DNA polymerase (Promega). Typically these reactions contained between 25 and 100 ng of template genomic DNA or equivalent cDNA, 5 μ l of 5 x reaction buffer, 1.5 mM MgCl₂, 200 μ M dNTPs, 0.2 μ M of each primer and 1 U of GoTaq DNA polymerase, combined into a 25 μ l reaction. DNA was denatured at 94°C for 2 min, followed by a variable number of cycles encompassing denaturation at 94°C for 30 s, primer annealing for 30 s (temperature determined using the melting temperature of primers), and a 30 s per kilobase extension

step at 72°C. The number of cycles varied depending on the abundance of the target as well as the efficiency of primer pairs.

2.3.2.2 High-fidelity polymerase chain reaction

When PCR was required to amplify cDNA for sequencing, or if long amplicons were generated, the LongAmp DNA polymerase (New England Biolabs) was used. The LongAmp polymerase exhibits 3'-5' exonuclease (proofreading) activity and thus ensured high-fidelity amplification. These reactions consisted up to 100 ng of template genomic DNA or equivalent cDNA, 5 µl of reaction buffer, 300 µM dNTPs, 0.4 µM of each primer, and 5 U of LongAmp DNA polymerase combined into a 25 µl reaction. Reactions were cycled as described for GoTaq above, however extension was performed at 65°C with a 1 min per kilobase extension time. For difficult to amplify templates or high GC-rich regions, 5% dimethyl sulfoxide (DMSO) was used in PCR reactions.

2.3.2.3 Agarose gel electrophoresis and gel extractions

PCR products were resolved by agarose gel electrophoresis (AGE). Molecular biology grade agarose (Amresco) was melted in 1 X TAE buffer [40 mM tris, 5.7% glacial acetic acid, 1 mM EDTA] and supplemented with SYBR Safe DNA stain (Invitrogen, Life Technologies). Samples were combined with 10 x loading buffer [4.4 mM Orange G, 40% glycerol] and separated by AGE alongside either 100 bp or 1 kb DNA markers (Invitrogen). DNA was visualised and photographed on a Gel Doc 2000 gel documentation system (Bio-Rad). PCR products destined for sequencing or for use as cDNA probes were isolated using the QIAquick gel extraction kit (QIAGEN) according to the manufacturer's protocol. Briefly, bands were excised using sterile scalpel blades, column purified and resuspended in 30 µl of H₂O. Extracts were stored at -20°C until required for downstream analysis.

2.3.2.4 First strand cDNA synthesis

First strand cDNA synthesis was performed using the Omniscript Reverse Transcriptase kit (QIAGEN) according to the manufacturer's instructions. RNA was first denatured at 65°C for 5 min to reduce secondary structure and then snap chilled on ice for a further 5 min. Reactions were assembled on ice and consisted of 1 µg of total RNA, 2 µl of reaction buffer, 500 µM dNTPs, 1 µM OligodT primer, 10 U of RNase inhibitor, and 4 U of Omniscript reverse transcriptase. Reactions were adjusted to 20 µl total volume with nuclease free H₂O mixed thoroughly and incubated at 37°C for 1 h. After extension, specimens were diluted 1 in 5 with nuclease free H₂O and stored at -80°C until required.

2.3.3 Rapid amplification of cDNA ends

Rapid amplification of cDNA ends (RACE) was performed using the GeneRacer kit (Invitrogen) to facilitate sequencing of full-length *CTGF* mRNA transcripts present in the pre-B ALL cell line PER-377. This method consists of three principal steps, preparing RNA for RACE amplification, reverse transcribing cDNA, and finally amplification by PCR.

2.3.3.1 Preparation of RACE-ready RNA

Ligation of a 44 base RNA oligonucleotide (5' GeneRacer Oligo) to the 5' end of decapped mRNA is the first step in the RACE protocol. The 5' GeneRacer oligo contains two priming sites that facilitate PCR amplification of the 5' end of transcripts in concert with a gene specific primer (GSP) after cDNA synthesis. Dephosphorylation and then decapping of 2 µg of DNase-treated total RNA was undertaken in accordance with the manufacturer's instructions without deviation. The 5' GeneRacer RNA oligonucleotide was then ligated to the 5' end of full length mRNA.

2.3.3.2 RACE cDNA synthesis

Full-length cDNA was reverse transcribed using SuperScript reverse transcriptase (RT) included in the GeneRacer kit and the GeneRacer OligodT primer which consists of 36 bases of complex sequence followed by 24 dT bases. This OligodT primer facilitated priming of first strand synthesis by annealing to polyadenylated mRNA and also incorporated known priming sites into the 3' end of all cDNA. Prior to reverse transcription, 2 µg of RACE-ready RNA, 0.5 mM dNTPs and 2.5 µM GeneRacer OligodT primer was combined in 3 µl H₂O and incubated at 65°C for 5 min to remove RNA secondary structure, then snap-chilled on ice for 2 min. The cDNA synthesis reaction consisted of 3 µl of denatured RACE-ready RNA, dNTPs and GeneRacer OligodT primer, combined in 20 µl with first strand reaction buffer, 5mM DTT, 40 U of RNase inhibitor and 200 U of SuperScript RT enzyme. Synthesis proceeded at 50°C for 1 h and this was followed by incubation at 70°C for 20 min to deactivate the RT enzyme. RNA was then digested by the addition of 2 U of RNase H at 37°C for 20 min.

2.3.3.3 Amplification of RACE clones by polymerase chain reaction

LongAmp DNA polymerase was used to amplify full-length race cDNA by nested PCR. For 5' RACE, the gene specific reverse primers used were CD_7R in the primary PCR, and CD_5R in the nested secondary PCR. For 3' RACE, the CD_1F forward primer was used in the primary PCR, followed by CD_1F_nested in the secondary PCR.

2.3.3.4 Band-stab polymerase chain reaction

The band stab technique (Bjourson & Cooper 1992), was used to isolate and re-amplify 5' and 3' RACE cDNA using LongAmp DNA polymerase. Bands of interest were inoculated into fresh PCR mix using sterile pipette tips and PCR amplification was performed using the same primer pair and thermocycling conditions as those used to amplify the original cDNA.

2.3.4 Quantitative real-time polymerase chain reaction

All reagents, assays, equipment and software were obtained from Applied Biosystems unless otherwise stated.

2.3.4.1 Quantitation of *CTGF* gene expression

Quantitative real-time PCR (qRT-PCR) was performed to measure expression of *CTGF* mRNA in cell lines and primary specimens. *CTGF* expression was measured using the on-demand assay Hs00170014_m1. Reactions were performed in 384 well plates on an ABI 7900HT Fast Real-Time PCR System, with the following thermal profile, 50°C for 2 min then 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. *CTGF* reactions were prepared by combining the following components per well, 1 µl of *CTGF* assay, 10 µl of universal PCR master mix, 8 µl of H₂O and 1 µl of first strand cDNA. *CTGF* gene expression was normalised using a custom *ACTB* assay comprising 10 µl of universal PCR master mix, 0.4 µl of 10 µM *ACTB* probe, 0.2 µl each of the *ACTB*_forward and *ACTB*_reverse primers (10 µM), 8.2 µl of H₂O and 1 µl of cDNA. Standards were prepared from cDNA generated from the PER-377 cell line. This cDNA was not diluted after cDNA synthesis and thus was approximately 5 times more concentrated than test specimens which were diluted 1 in 5 immediately after cDNA synthesis. Standards were prepared by serial dilution and assayed in duplicate to control for pipette error, while test specimens were assayed in triplicate. The relative expression for each assay was calculated using the standard curves.

2.3.4.2 *CTGF* gene copy number analysis

CTGF Gene copy number analysis was performed using qRT-PCR. The *CTGF* gene copy number assay Hs00517234_cn targets exon 5 within the coding region. This assay was run in duplex with the Taqman reference copy number assay for the *RPPHI* locus. Assays were performed in 384 well plates on an ABI 7900HT Fast Real-Time PCR System with the following thermal profile, 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. *CTGF* and *RPPHI* copy number assays were run in

duplex reactions prepared by combining the following components per well, 0.5 µl *CTGF* assay, 0.5 µl *RPPHI* reference assay, 5 µl of universal PCR master mix, 2 µl of H₂O and 2 µl of genomic DNA at 5 ng/µl. Samples were assayed using 4 technical replicates. At the completion of the assay, data was imported into the Copy Caller software and analysed to determine *CTGF* gene copy number normalised to *RPPHI*.

2.3.5 Northern blotting of RNA

2.3.5.1 RNA electrophoresis and membrane transfer

Cytoplasmic RNA was isolated according to the method outlined in section 2.3.1.3, and 10 µg of cytoplasmic RNA was combined with 1 µl of 20 x northern running buffer (NRB) [122 mM Na₂HPO₄, 78 mM NaH₂PO₄] and 1.5 µl of deionised 8.8 M glyoxal in a final volume of 20 µl. Samples were denatured at 50°C for 1 h and then chilled on ice for 20 min. Samples were combined with 5 µl of sample loading buffer [50% glycerol, 0.02% bromophenol blue, 1 x NRB] and together with an RNA ladder (Invitrogen, Life Technologies), were resolved on 0.8% agarose in 1 x NRB. Gels were run slowly at 50V, at 4°C for 3.5 h. RNA was stained with a solution of 0.8 mg/ml ethidium bromide in 1 x NRB for 30 min, then de-stained in fresh NRB for a further 30 min. RNA was transferred to a neutral magna membrane (Osmonics) by capillary transfer in 20 x SSC [3 M NaCl, 300 mM Na Citrate, pH 7.0] overnight and cross linked to the membrane while still damp with 150 mJoules UV using a GS Gene Linker (Bio-Rad).

2.3.5.2 Northern blot probe synthesis

Hybridisation probes were generated by PCR amplification using GoTaq DNA polymerase. The template cDNA used to amplify these probes was reverse transcribed from RNA isolated from an airway fibroblast cell line (HFL-1). This RNA sample was a kind gift from Mr Bradley Shelton, Lung Institute of Western Australia, (Shenton Park, Australia). Two probes were generated targeting *CTGF* mRNA. The first probe designated as the Central probe, hybridised *CTGF* between exons 3 and 5 and was

amplified with the primer pair NB_central_for and NB_central_rev. The second probe hybridized *CTGF* mRNA in the 3'UTR and was amplified with the primer pair NB_3pr_for and NB_3pr_rev. To provide a loading control, a probe hybridizing *ATP5G3* was amplified using the primer pair ATP5G3_for and ATP5G3_rev.

2.3.5.3 Radiolabelling of cDNA probes

After gel extraction, 30 ng of purified cDNA probe was labelled with 50 μ Ci of 32 P dCTP radionuclide (Perkin Elmer) using the Rediprime II Random prime labelling system (GE Healthcare) according to the manufacturers instructions.

2.3.5.4 Northern blot probe hybridisation

Directly after cross linking, the RNA-bound membrane was soaked in 2 x SSC for 5 min at room temperature, then placed between two sheets of nylon mesh and added to a 50 cm hybridisation bottle together with 15 ml of hybridization buffer [5% dextran sulphate, 20% Denharts Solution, 0.5 mg/ml salmon sperm DNA, 3% SSC, 0.1% SDS] pre-heated to 50°C. Hybridisation bottles were incubated at 50°C for 1 h in a model 400 hybridisation incubator (Robbins Scientific). 32 P-labelled probe was denatured by boiling for 5 min then snap chilled on ice for 10 min. Denatured probe was then added to the hybridisation bottle and incubated in the rolling incubator overnight at 50°C. The next morning, the membrane was washed with four successive and increasingly stringent 15 min washes, comprised of 2 x SSC, 2 x SSC/0.1% SDS, 0.5 x SSC/1% SDS, 0.2 x SSC/0.1% SDS respectively. The membrane was then incubated with autoradiograph film at -80°C. Membranes were stripped for re-probing by pouring a boiling solution of 0.1 x SSC/0.1% SDS over membranes and allowing them to cool to room temperature. Densitometry analysis was performed using the ImageJ program (NIH).

2.3.6 Southern blotting

2.3.6.1 Restriction digest of genomic DNA

Genomic DNA from pre-B ALL cell lines or PBMCs was concentrated by ethanol/sodium acetate precipitation. Briefly, DNA solutions were combined with 2.5 vol of absolute EtOH and 0.1 vol of 3 M sodium acetate pH 5.2 and stored overnight at -20°C. The following day, samples were centrifuged at 15,000 x g for 15 min at 4°C. Following removal of the supernatant, pellets were washed with 70% EtOH and briefly air-dried. Once almost dry, pellets were resuspended in 50 µl nuclease-free H₂O. Samples were heated at 65°C for 30 min to ensure complete resuspension, and then lightly vortexed to avoid excessive pipetting which may result in shearing of full-length DNA. Samples were quantitated and 15 µg of gDNA was digested with 15 U of either Psi I (New England Biolabs) or Hind III (Promega) in 150 µl digests overnight at 37°C. The following day, restriction enzymes were inactivated by incubating samples at 65°C for 15 min and samples were concentrated by EtOH/sodium acetate precipitation as described above and resuspended in 15 µl of nuclease-free H₂O.

2.3.6.2 Electrophoresis and transfer of digested genomic DNA

Samples were separated on 0.8% agarose in TAE overnight at 40 V. The following day, the gel was stained with 1 x SYBR safe DNA stain (Invitrogen, Life technologies) in TAE buffer for 30 min and photographed under UV illumination to confirm adequate migration of digested DNA. The gel was then soaked in 300 ml of denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 40 min at room temperature, then rinsed twice with ddH₂O (Baxter) and soaked in two changes of neutralizing solution (0.5 M tris pH 7.4, 1.5 M NaCl) for a total of 1 h. DNA was then transferred to a neutral Magna membrane (Osmonics) by capillary transfer with 10 x SSC overnight. Following transfer, lanes were marked with a soft pencil and the membrane was soaked in 6 x SSC for 5 min. DNA was cross linked to the membrane while still damp with 150 mJoules UV using a GS Gene Linker (Bio-Rad).

2.3.6.3 Southern blot probe synthesis

A hybridisation probe for genomic Southern blotting was amplified from PBMC genomic DNA using LongAmp DNA polymerase. This DNA probe spanned the *CTGF* gene from -1024 to +3418 and was amplified with the primer pair SB_CTGF_for and SB_CTGF_rev. A 1050 bp probe which hybridized the *IL2RB* gene was a kind gift from Dr M. Hatakeyama, Osaka (Hatakeyama et al. 1989).

2.3.6.4 Radiolabelling of probes

After gel extraction, 30 ng of gel purified probe was labelled with 50 μ Ci of 32 P dCTP radionuclide (Perkin Elmer) using the Rediprime II Random prime labelling system (GE Healthcare) according to the manufacturers instructions.

2.3.6.5 Southern probe hybridisation

Membrane and mesh was pre-soaked in 2 x SSC for 5 min at room temp, then rolled and added to a hybridisation bottle together with 15 ml of pre-heated hybridization buffer [2 x SSPE, 7% SDS, 0.5% Blotto, 0.5 mg/ml salmon sperm DNA] and incubated at 65°C for 1 h. 32 P-labelled DNA probe was denatured by boiling for 5 min then snap chilled on ice for 10 min. Denatured probe was then added to the hybridisation bottle and incubated overnight at 65°C. The next morning, the membrane was washed with four successive 15 min washes comprised of 2 x SSC/0.1% SDS, 2 x SSC/0.1% SDS, 1 x SSC/0.1% SDS, 0.5 x SSC/0.1% SDS. The membrane was then incubated with autoradiograph film at -80°C. Membranes were stripped for re-probing by pouring a boiling solution of 0.1 x SSC/0.1% SDS over membranes and allowing them to cool to room temperature.

2.3.6.6 Electrophoresis and transfer of RACE-PCR products

Band-stab PCR products were examined by Southern blotting with a *CTGF*-specific probe. PCR products were separated on 1.5% agarose in TAE and stained with SYBR

Safe DNA stain. PCR products were then transferred to a positively charged N⁺ Hybond membrane (Amersham, GE Healthcare) by capillary transfer with 0.4 M NaOH overnight. Following transfer, lanes were marked with a soft pencil and the damp membrane was cross linked with 150 mJoules UV using a GS Gene Linker (Bio-Rad). The probe used to hybridise band-stab PCR products was amplified from cloned full-length *CTGF* cDNA with GoTaq DNA polymerase using the CD_1F and CD_5R primers, then radiolabelled and hybridised to membrane-bound RACE PCR products according to the method outlined above in section 2.3.6.5.

2.3.7 DNA cloning

2.3.7.1 Preparation of chemically competent cells

TOP10 *E. coli* were a gift from Mrs Marie Scobie. Single colony isolation was performed by streak isolating on LB agar plates grown at 37°C overnight. A 10 ml culture of LB broth [10 g Bacto-tryptone (BD Bioscience), 5 g yeast extract (BD Bioscience), 170 mM NaCl] was inoculated with a single colony and grown overnight at 30°C with moderate shaking. The following day, 99 ml of LB agar was pre-warmed to 30°C, and then inoculated with 1 ml of overnight culture. This culture was grown at 30°C with moderate shaking and the absorbency at 600 nm was measured regularly using a BioPhotometer spectrophotometer (Eppendorf). Once the OD₆₀₀ reached 0.5, cultures were immediately transferred to a 4°C cold-room and chilled on ice for 10 min. All remaining steps were where possible performed at 4°C using pre-chilled glassware and instruments. Cells were collected by centrifugation at 3000 x g for 10 min. The supernatant was discarded and cells were gently resuspended in 20 ml of ice-cold, 50 mM CaCl₂. Cells were collected by centrifugation at 3000 x g for 5 min and then resuspended gently in ice-cold 75 mM CaCl₂/25% glycerol. This suspension was added to pre-chilled, sterile 1.5 ml tubes in 100 µl aliquots, then snap frozen by immersion in liquid nitrogen. Competent cells were stored at -80°C until use.

2.3.7.2 Cloning into the pGEM-T Easy vector

PCR products were cloned using the pGEM-T Easy TA cloning system (Promega) according to the manufacturer's protocol. Inserts were invariably generated using the LongAmp DNA polymerase. While this enzyme exhibits 3'-5' exonuclease activity, the majority of PCR products contain dA overhangs at the 3' end allowing cloning into TA vectors such as pGEM-T Easy. After gel purification, PCR products were quantitated and 5 to 10 ng of insert was combined with ligation buffer, pGEM-T Easy vector (50 ng) and 3 U of T4 DNA ligase in a 10 µl reaction. Ligations were incubated overnight at 4°C.

2.3.7.3 Transformation of chemically competent cells

Chemically competent cells were thawed on ice for 15 min. After thawing, 5 µl of ligation reaction was added to competent cells followed by a further 15 min incubation on ice. Cells were then subjected to heat shock transformation by submersion in a 42°C water bath for 1 min, then transferred to ice immediately and incubated for 15 min. To maximise transformation, this process of heat shock and recovery was repeated once more. After the second 15 min recovery, 250 µl of SOC medium [10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose] was added to cells and tubes were incubated for 1 h at 37°C with moderate shaking. LB agar plates were prepared containing 100 µg/ml ampicillin. Thirty mins prior to spreading transformants, plates were supplemented with 50 µl of 200 mM IPTG and 20 µl of 50 mg/ml X-Gal and placed at 37°C. At the completion of the 1 h incubation, 150 µl of transformed cells were spread on the prepared selective agar plates and incubated overnight at 37°C.

2.3.7.4 Transformant screening and plasmid DNA extraction

The following day, blue-white screening revealed clones carrying DNA inserts and these clones were used to inoculate LB broth supplemented with 100 µg/ml ampicillin. These 5 ml cultures were grown overnight at 37°C with vigorous shaking. The following day, glycerol stocks were prepared from each culture by combining 100 µl of culture with

100 μ l of 80% glycerol. These stocks were frozen immediately at -80°C . The remainder of the overnight culture was centrifuged in preparation for plasmid DNA extraction using the QIAprep Spin Miniprep kit (QIAGEN) according to the manufacturer's instructions. Purified plasmid DNA was eluted in 50 μ l of elution buffer and stored at -20°C .

2.3.7.5 Isolation of bacterial artificial chromosome DNA

The Human TilePath Library clone RP11-69I8 was purchased from the Murdoch Research Institute (Melbourne, Australia). This bacterial artificial chromosome (BAC) contained a fragment of chromosome 6 that included the *CTGF* locus. A single colony was streak isolated on an LB agar plate supplemented with 12.5 $\mu\text{g/ml}$ chloramphenicol. This colony was used to inoculate 50 ml LB broth containing the same concentration of chloramphenicol and this culture was grown overnight at 37°C with shaking. The following day, cells were collected by centrifugation at $3000 \times g$ for 10 min at 4°C . Cells were resuspended in 50 ml ice-cold STE buffer [10 mM Tris-Cl (pH 8.0), 0.1 M NaCl, 1 mM EDTA (pH 8.0)]. Cell were then collected by centrifugation again and resuspended in 1 ml of ice-cold alkaline lysis solution I [50 mM glucose, 25 mM Tris-Cl (pH 8.0), 10 mM EDTA (pH 8.0)] and transferred to five 1.5 ml tubes (200 μ l each) and placed on ice. After 5 min, 400 μ l of alkaline lysis buffer II [0.2 M NaOH, 1% w/v SDS] was added to tubes and mixed by gentle inversion several times. Next, 300 μ l of ice-cold alkaline lysis buffer III [3 M potassium, 5 M acetate] was added and tubes were mixed gently by inversion and incubated on ice for 5 min. Debris was cleared by centrifugation at $15,000 \times g$ for 5 min at 4°C and supernatants were decanted into fresh 2 ml tubes. At room temperature, 900 μ l of isopropanol was added and the tube was mixed again by gentle inversion. Precipitated nucleic acids were then collected by centrifugation at $15,000 \times g$ for 5 min at room temperature. Pellets were washed once with 70% EtOH and briefly air dried, then resuspended in 20 μ l of TE buffer. Tubes were combined and stored at -20°C .

2.3.8 DNA Sequencing

DNA sequencing was carried out using BigDye version 3.1 (Applied Biosystems). Sequencing reactions were performed in 96 well plates suitable for the 3130x Genetic Analyser sequencing platform (Applied Biosystems). Briefly, 75 ng of plasmid DNA or 100 ng of gel extracted PCR products were combined with 3.5 μ l of 5 x BigDye reaction buffer, 9.3 μ l of ultrapure H₂O (Baxter), 3.2 μ l of 1 μ M sequencing primer and 1 μ l of BigDye v3.1. Samples were denatured at 96°C for 1 min, followed by 25 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 4 min. Inserts cloned into the pGEM-T Easy vector were sequenced using M13 forward and reverse primers listed in Table A2.

2.3.8.1 BigDye reaction purification and analysis

BigDye reaction products were purified *in situ* by EtOH/sodium acetate precipitation. To each well, 2 μ l of 3 M sodium acetate (pH 5.2) and 50 μ l of absolute EtOH was added. The plate was resealed, mixed by vortexing, and spun for 30 s at 1000 x g to collect the solution to the bottom of wells. The plate was then incubated at room temperature in the dark for 2 h. BigDye products were then pelleted by centrifugation at 15,000 x g for 30 min at 4°C. The supernatant was removed by inversion and pellets were washed using 70% EtOH and centrifuged again at 15,000 x g for 10 min at 4°C. Supernatants were removed by inversion and the pellets were dried by inverting the plate on lint-free paper towel and centrifuging at 800 x g for 1 min.

After briefly air drying pellets, 10 μ l of HiDi formamide (Applied Biosystems) was then added to each well and the pellets were denatured by two consecutive rounds of 96°C for 2 min followed by incubation on ice for 2 min. Plates were then loaded onto a 3130x Genetic Analyzer (Applied Biosystems) for capillary sequencing. Sequencing data generated from BigDye sequencing was analysed using the KB Base Caller software (Applied Biosystems).

2.3.8.2 *CTGF* promoter and 3'UTR sequencing

The promoter and 3' UTR regions of the *CTGF* locus were amplified using the high-fidelity LongAmp DNA polymerase. The PR_3F and PR_1R primers were used to generate a PCR amplicon spanning the *CTGF* proximal promoter and this product was gel extracted and sequenced using the primers; PR_3F, PR_3R, PR_2F, PR_2R, PR_1F and PR_1R. The 3'UTR was amplified in 3 separate PCR amplicons using primer pairs; CD_7F and CD_7R, CD_8F and CD_8R, CD_9F and CD_9R and these three amplicons were gel purified and sequenced in forward and reverse orientation using these primers.

2.3.9 Analysis of DNA methylation

2.3.9.1 Conversion of DNA by sodium bisulfite

To investigate DNA methylation at the *CTGF* locus a number of techniques were performed that relied upon effective bisulfite conversion of non-methylated cytosines in the genome by sodium bisulfite. The EpiTect Bisulfite Kit (QIAGEN) was used according to the manufacturer's instructions to convert genomic DNA from cell lines and primary specimens, as well as BAC DNA which served as a control in methylation specific PCRs (outlined below).

2.3.9.2 BAC DNA methylation

The human BAC Tile Path clone RP11-69I8 containing the entire *CTGF* coding region was methylated using the CpG methyltransferase M.SssI (New England Biolabs) to serve as a positive control in methylation specific PCRs. Reactions consisted of 1 µg of BAC DNA, 160 µM S-adenosylmethionine (SAM), reaction buffer and 4 U of methyltransferase in 20 µl. Reactions were performed at 37°C for 3 h. The enzyme was then deactivated by incubating at 65°C for 20 min. Converted BAC DNA was diluted to 1000 µl with H₂O and stored at -20°C.

2.3.9.3 Methylation specific PCR

Bisulfite converted DNA was subjected to PCR amplification using primers specific for either methylated or unmethylated priming sites. These primer pairs were designated “methylated-site PCR” (MSP_for, MSP_rev) or “unmethylated site PCR” (USP_for, USP_rev), and have been used successfully in other studies (Chiba et al. 2005). Methylation specific PCR (msPCR) reactions were performed with the GoTaq DNA polymerase as outlined in section 2.3.2.1, using 100 ng of bisulfite-treated genomic DNA. Bisulfite-treated BAC DNA (1 ng) served as a positive control for the USP reaction, while M.SssI-methylated, bisulfite-treated BAC DNA served as a positive control for the MSP reaction.

2.3.9.4 Bisulfite Sequencing

Three separate PCR amplicons spanning the *CTGF* CpG island were amplified using LongAmp DNA polymerase. These amplicons, described previously (Kikuchi et al. 2007) were generated from 100 ng of bisulfite treated genomic DNA. Amplicon 1 was amplified with the primer pair Bis1_for and Bis1_rev. Bisulfite amplicons 2 and 3 were amplified in nested PCR reactions using the following primers; Bis2-primary_for and Bis2_primary_rev, Bis2_secondary_for and Bis2_secondary_rev, Bis3-primary_for and Bis3_primary_rev, Bis3_secondary_for and Bis3_secondary_rev. PCR products were gel extracted using the QIAquick Gel Extraction Kit (QIAGEN) and eluted in 50 µl of H₂O. Purified bisulfite amplicons were cloned into the pGEM-T Easy vector according to the method outlined in section 2.3.7.2 and plasmid DNA was purified for sequencing from a selection of transformants.

Sequencing of cloned bisulfite amplicons was performed according to the method described in section 2.3.8, however bisulfite-converted DNA has a significantly lower melting temperature, and thus sequencing of bisulfite clones required a different thermal profile for BigDye reactions. After denaturation at 96°C for 1 min, samples were subjected to 24 cycles of 96°C for 10 s followed by extension at 50°C for 4 min. Bisulfite clones were analysed for effective bisulfite conversion and incidence of CpG methylation using BiQ Analyzer software (Bock et al. 2005).

2.3.10 Immunological detection of CTGF protein

2.3.10.1 Isolation of total cellular protein

Total protein was isolated from cell lines for analysis by western blotting. Cells were pelleted by centrifugation and washed twice with ice-cold PBS. Cell pellets were resuspended in one pellet volume of ice-cold PBS and then lysed using 50 μ l of ice-cold protein lysis buffer [2 mM ethylenediaminetetraacetic acid (EDTA), 10 mM iodoacetamide, 25 μ g/ml p-nitrophenylguanidinobenzoate, 10 μ g/ml Leupeptin, 10 μ g/ml Aprotinin and 0.5% Triton-X] per 10^6 cells, and incubated on ice for 30 min. Samples were centrifuged for 20 min at 15,000 x g at 4°C and supernatants were transferred to fresh 1.5 ml tubes. Protein samples were stored at -20°C or used immediately in downstream assays.

2.3.10.2 Protein estimation

Protein concentration in cell line extracts was determined using a commercial Bradford assay (Bradford 1976). Protein concentration was re-established each time samples were thawed prior to western blotting. Bio-Rad Protein Assay (Bio-Rad) dye was diluted 1 in 4 with H₂O and then added to a 96 well micro-titre plate (Becton Dickinson) at 200 μ l per well. A stock solution of 10 mg/ml BSA was serially diluted to provide a standard curve within the range of 5 mg/ml to 10 μ g/ml. Samples were diluted 1 in 10 and together with standards, were added to the plate at 10 μ l per well. Standards were assayed in duplicate, while protein samples were assayed in triplicate. After addition of samples and standards, the plate was incubated at room temperature for 5 min, and the absorbance was measured at 595 nm on a Victor² 1420 multilabel counter (Wallac, Perkin Elmer). Protein concentration in cell line extracts was calculated using the standard curve.

2.3.10.3 Sample preparation

Prior to SDS polyacrylamide gel-electrophoresis (SDS-PAGE), 75 µg of total protein was combined with 3.3 x sample buffer [0.2 M Tris-HCl, 40% glycerol, 8.3 mM dithiothreitol, 4% w/v SDS, 0.008% bromophenol blue and 6% 2-mercaptoethanol] and placed in boiling water for 5 min. Samples were chilled on ice for 5 min then loaded onto the gel or frozen immediately at -20°C until required. If frozen, samples were denatured by boiling again prior to electrophoresis.

2.3.10.4 Heparin-affinity enrichment of conditioned culture medium

Cells were seeded in fresh culture medium at a density of 10^6 cells per ml on 24 well culture plates (Nunc, Thermo Fisher Scientific), and grown for 24 h prior to isolation of conditioned medium (CM). Cells and CM were collected and centrifuged at 800 x g for 5 min and supernatants comprising the CM was then transferred to fresh tubes and chilled on ice. After 5 min, 5 ml of ice-cold PBS was added to fresh 10 ml tubes together with 150 µl of heparin agarose beads. 5 ml of chilled CM was added to these tubes and incubated for 24 h at 4°C with constant gentle mixing. After 24 h, beads were collected by centrifugation at 800 x g for 5 min at 4°C then allowed to stand for 2 min at 4°C. Supernatant was carefully removed by vacuum aspiration and beads were washed twice by this method with ice-cold PBS. After the second wash, beads were resuspended in 1 ml ice-cold PBS and transferred to 1.5 ml tubes. Beads were collected by centrifugation at 800 x g for 3 min at 4°C and the supernatant was carefully aspirated. Bound protein was eluted by adding 200 µl of protein loading buffer and denatured in boiling water for 5 min. Tubes were centrifuged while still hot at 10,800 x g for 3 min to collect beads. The supernatant containing eluted protein was removed to fresh 1.5 ml tubes and either subjected immediately to SDS-PAGE or frozen at -80°C until required.

2.3.10.5 Western blotting

Denatured protein samples were separated by 15% SDS-PAGE together with Precision Plus Protein Dual Colour Standards (Bio-Rad), before being transferred to Hybond-C

Super membrane (AP Biotech) in transfer buffer [25 mM tris, 190 mM glycine, 20% methanol, pH 8.2]. Membranes were blocked with 5% skim milk powder (SMP) in wash buffer [50 mM tris, 150 mM NaCl, 0.05% Tween] at room temperature for 1 h and incubated with primary antibody in 5% SMP in wash buffer overnight at 4°C. Primary antibody was removed and the membrane rinsed twice with wash buffer, then washed three times for 15 min with wash buffer. Membranes were incubated with either horse radish peroxidase (HRP) or biotin conjugated secondary antibodies, in 5% SMP in wash buffer for 1 h at room temperature. After three 15 min washes with wash buffer, membranes incubated with biotinylated secondary antibodies were incubated with a streptavidin conjugated HRP (Dako) for 5 min then re-washed three times in wash buffer. Visualisation of bands was performed by enhanced chemiluminescence using the Immobilon HRP substrate (Millipore).

2.3.10.6 Antibodies

The two CTGF antibodies used in this study included the polyclonal L20 goat anti-CTGF antibody (Santa Cruz Biotechnology), which targets amino acids 150-200 of the CTGF protein, and the polyclonal Rabbit-A rabbit anti-CTGF antibody, which targets amino acids 247-260 of the CTGF protein. This antibody was a generous gift from Prof David Brigstock (Brigstock et al. 1997). The ACTN05 monoclonal mouse anti-ACTIN antibody (Neomarkers) was used to detect ACTIN which served as a loading control. Secondary, HRP-conjugated antibodies included the anti-rabbit Na9340V antibody (GE Healthcare) and the anti-mouse Na 9310V antibody (GE Healthcare). For detection of primary goat antibodies, the biotinylated EO466 rabbit anti-goat antibody (Dako) was used followed by detection with streptavidin-conjugated HRP (Dako).

2.3.10.7 Membrane stripping protocol

Where it was necessary to strip and re-probe membranes, they were incubated in stripping buffer [50 mM Tris (pH 6.8), 2% SDS, 100 mM 2-mercaptoethanol] at 50°C for 30 min with gently shaking, then washed three times in western wash buffer for 15

min at room temperature. These membranes were then blocked and probed as previously described.

Table A2.1 List of suppliers

Amresco	- Solon, Ohio, USA
AP Biotech	- Little Chalfont, Buckinghamshire, UK
Baxter	- Old Toongabbie, New South Wales, Australia
BD Bioscience	- San Jose, California, USA
Becton Dickinson	- Franklin Lakes, New Jersey, USA
Bio-Rad	- Hercules, California, USA
Dako	- DK-2600, Glostrup, Denmark
GE Healthcare	- Little Chalfont, Buckinghamshire, UK
Geospiza	- Seattle, Washington, USA
GraphPad Software	- De La Playa, California, USA.
Ibis Biosciences	- Carlsbad, California, USA
Life Technologies	- Carlsbad, California, USA
Microsoft	- Redmond, Washington, USA
Milipore	- Bedford, Massachusetts, USA
MJ Research	- Waltham, Massachusetts, USA
MP Biomedicals, LLC	- Santa Ana, California, USA
Neomarkers	- Fremont, California, USA
New England Biolabs	- Ipswich, Massachusetts, USA
Nycomed	- Asker, Oslo, Norway
Osmonics	- Westborough, Massachusetts, USA
Perkin Elmer	- Waltham, Massachusetts, USA
QIAGEN	- Hilden, Germany
Robbins Scientific	- Sunny Vale, California, USA
Santa Cruz Biotechnology	- Santa Cruz, California, USA
Sigma Aldrich	- St Louis, Missouri, USA
Thermo Fisher Scientific	- Waltham, Massachusetts, USA

Table A2.2 Primer and oligonucleotide sequences

<u>primer name</u>	<u>sequence 5'-3'</u>
ACTB_forward	GGCACCCAGCACAATGAAG
ACTB_probe	VIC_TCAAGATCATTGCTCCTCCTGAGCGC_TAMRA
ACTB_reverse	GCCGATCCACACGGAGTACT
ATP5G3_for	TAATCCAAAGGGAGTTTCAGAC
ATP5G3_rev	AATCAAGAAAGCAACCATCAAAC
Bis1_for	GTAGGAAGGTGGGGAGGAA
Bis1_rev	CACTAACTATCTCCTCTCAAC
Bis2_primary_for	GGAATGTTGAGTGTTAAGGGGTTAGGATTA
Bis2_primary_rev	ATCAAACATTAAAACACTCTCACATCCAAA
Bis2_secondary_for	TTGAGAGGAGATAGTTAGTG
Bis2_secondary_rev	AACAAAATAAACCCCTTATAC
Bis3_primary_for	GGTTGTTAGGGAGGGATT
Bis3_primary_rev	TCCATACTACACAAAACATACAACC
Bis3_secondary_for	GTATAAGGGTTTATTTTGTATTT
Bis3_secondary_rev	CACTAATACTTACA ACTACTCTA
CD_1F	AGTGCGACTCCACCCTCCA
CD_1F_nested	CCTCCAGCTCGACGGCAG
CD_4F	AAGGACCAAACCGTGGTTGG
CD_5F	GACCTGGAAGAGAACATTAAGG
CD_5R	TACTCCACAGAATTTAGCTCGG
CD_6R	TTAATGTCTCTCACTCTCTGGC
CD_7F	TGTACTACAGGAAGATGTACGG
CD_7R	AGAATGTCAGAGCTGAGTCTGC
CD_8F	TTAGCGTGCTCACTGACCTG
CD_8R	TTCTACCTAGAAATCAGCC
CD_9F	TGCTCAGATAGAATGACAGTCC
CD_9R	TCAATGTGGTGTGTATGCCTGC
GeneRacer 3'	GCTGTCAACGATACGCTACGTAACG
GeneRacer 3' nested	CGCTACGTAACGGCATGACAGTG

Table A2.2 (continued)

<u>primer name</u>	<u>sequence 5'-3'</u>
GeneRacer 5'	CGACTGGAGCACGAGGACTGA
GeneRacer 5' nested	GGACTGACATGGACTGAAGGAGTA
GeneRacer OligodT	GCTGTCAACGATACGCTACGTAACGGCATGACAGTGT(24)
GeneRacer RNA oligo	CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAGUAGAAA
M13_for	GTAAAACGACGGCCAGTG
M13_rev	CAGGAAACAGCTATGAC
MSP_for	TCGTTTCGGTCGATAGTTTC
MSP_rev	CGAAACCCATACTAACGACG
NB_3pr_for	AGGGTACCAGCAGAAAGGTTAGTA
NB_3pr_rev	AGAAATCAGCCTGCCAAGGACT
NB_central_for	GAGTGGGTGTGTGACGAGCCCAAGG
NB_central_rev	ATGTCTCCGTACATCTTCCTGTAGT
PR_1F	AGTGGACAGAACAGGGCA
PR_1R	GCGGCTGCCGTCGAGCTG
PR_2F	CAGGTAGGCATCTTGAG
PR_2R	CACTGGCTGTCTCCTC
PR_3F	ACAACAGGGTACACTGCTC
PR_3R	TGATTTGCGTTTTAGAGGC
SB_CTGF_for	CCTTTGCTGGCCATTCACACTATTG
SB_CTGF_rev	ACCCCTACTAGATTCAACAGCTGCT
USP_for	TTGTTTTGGTTGATAGTTTT
USP_rev	CAAAACCCATACTAACAACA

Chapter 3

Gene Expression and Clinical Correlates of *CTGF* Expression in Childhood Pre-B ALL

Chapter 3

Gene Expression and Clinical Correlates of *CTGF* Expression in Childhood Pre-B ALL

3.1 Introduction

Oncogenic events contributing to leukaemogenesis include the activation of proto-oncogenes that promote dysregulation of cell cycle controls, apoptosis, and differentiation (Pui 2006). Aberrant *CTGF* gene expression occurs in around 75% of pre-B ALL cases (Vorwerk et al. 2000; Boag et al. 2007; Sala-Torra et al. 2007), however the mechanism underlying this dysregulated expression is not known. Activation of *CTGF* transcription in other cell types has been attributed to numerous soluble factors such as TGFB1 (Arnott et al. 2008), SRF (Muehlich et al. 2007) and MMP3 (Eguchi et al. 2008), as well as changes in the microenvironment such as hypoxia (Hong et al. 2006), mechanical stretch (Nishida et al. 2008) and high glucose (Wang et al. 2008). Regulation of the *CTGF* promoter is complex and is heavily dependent on cell type and tissue context (Cicha & Goppelt-Struebe 2009). Control of *CTGF* expression and its role in the cellular biology of lymphocyte precursors remains ill-defined as there has only been one study to date that has identified *CTGF* gene expression in normal haemopoietic precursor cells (Wagner et al. 2007), and no detailed characterisation of *CTGF* gene expression in haemopoietic cells has been reported.

Analysis of global gene expression data may provide some clues. Identifying transcriptional programs associated with *CTGF* gene expression by discerning genes that are highly correlated with *CTGF*, could point to the involvement of common transcription factors or regulatory pathways (Eisen et al. 1998). Global gene expression data gathered from a cohort of 73 paediatric patients diagnosed with pre-B ALL has been recently generated by this laboratory (Boag et al. 2006; Boag et al. 2007). This has provided a valuable resource to investigate whether any genes display patterns of expression similar to *CTGF*. This approach was expected to shed light on putative DNA

regulatory elements or transcription factors contributing to *CTGF* deregulation, through analysis of the regulatory sequences of co-regulated genes in pre-B ALL.

While its role in the leukaemogenesis of pre-B ALL is unclear, several studies have highlighted associations between *CTGF* gene expression and patient outcome in high-risk pre-B ALL. Sala-Torra and colleagues demonstrated a significant difference in EFS in adult ALL patients stratified based upon *CTGF* mRNA expression levels exhibited by their leukaemic cells. Specifically, patients whose leukaemic blasts expressed *CTGF* at low levels showed 5yr EFS of 58%, in contrast to 12% for those with high *CTGF* expression (Sala-Torra et al. 2007). Clearly, high *CTGF* expression was associated with a profoundly poor outcome for these patients. In relation to childhood ALL, a recent COG study investigating methods of improving stratification of high-risk pre-B ALL patients by microarray, successfully used *CTGF* expression in a multi-gene classifier model to predict the response of patients to therapy (Kang et al. 2010). As in adult ALL, high *CTGF* expression in this paediatric cohort was associated with a poor outcome and shorter period of RFS. These studies support the hypothesis that *CTGF* protein may play an important role in the pathobiology of ALL. *CTGF* expression has been associated with a metastatic and aggressive phenotype in other cancers (Shimo et al. 2006; Aikawa et al. 2006; Kang et al. 2003), however despite clear association with patient outcomes, no progress has been made in identifying biological features associated with *CTGF* gene expression in pre-B ALL. Analysis of clinical features associated with *CTGF* expression may point to a role for this pleiotropic protein in leukaemogenesis or maintenance of leukaemic cells.

In this study, analysis of *CTGF* gene expression in primary paediatric pre-B ALL specimens was performed to establish gene expression and clinical correlates with *CTGF*. Specifically, the aims of this chapter were to 1) validate *CTGF* expression data generated by microarray in primary specimens using qRT-PCR, 2) determine genes correlated with *CTGF* mRNA levels in three independent pre-B ALL cohorts, and compare their promoters with respect to common regulatory motifs, and 3) examine

whether *CTGF* expression is linked to diagnostic features or prognosis in paediatric pre-B ALL.

3.2 Results

3.2.1 The PMH cohort of paediatric pre-B ALL

The global gene expression profiles of primary pre-B ALL specimens was measured by Boag and colleagues using Affymetrix microarray technology (Boag et al. 2006; Boag et al. 2007). This group of specimens, named the PMH cohort, comprised 73 paediatric pre-B ALL specimens from patients treated on COG protocols (Gaynon et al. 2000). These were obtained either at the time of diagnosis; n=60, or at relapse; n=13. The critical clinical and biological features of this cohort are detailed in Table 3.1.

Prior to analysing the global gene expression data from the PMH cohort (Boag et al. 2006; Boag et al. 2007), *CTGF* expression levels measured by microarray were validated using 17 primary pre-B ALL specimens from this cohort. Also included were 3 T-ALL specimens for comparison. The details of these specimens are outlined in Table 3.2. These specimens, also used in later parts of the current study (Chapters 4 & 6) were assayed for *CTGF* expression by quantitative real-time PCR (qRT-PCR). RNA from these 20 primary bone marrow samples was extracted from cryopreserved material that was matched with those used for the previous microarray study (Boag et al. 2007). These specimens represented a range of cytogenetic subtypes, determined using a previously published 20 gene classifier model for cytogenetic prediction (Hoffmann et al. 2006). These were *MLL* (n=1), *BCR-ABL* (n=4), hyperdiploidy (n=2), *TEL-AML* (n=2), *TCF3-PBX1* (n=1) and 10 cases where known cytogenetic abnormalities could not be detected (Other). These latter specimens were all classed as low/neg for *CTGF* mRNA by microarray as they exhibited RMA normalised probeset signals of <50 (see section 2.2.1.1).

Table 3.1 Clinical and biological features of pre-B ALL patients in the PMH cohort

	Number	%		Number	%
Age			CNS involvement		
1-10 years	53	73	Not known	4	5.5
>10 years	20	27	Present	1	1.5
Gender			Absent	68	93
Male	41	56	Spleen		
Female	32	44	Normal	32	44
WBC[§]			Enlarged (BU) ‡	37	51
<50 x 10 ⁹ /l	67	92	Enlarged (AU) ‡	4	5
>50 x 10 ⁹ /l	6	8	Lymph nodes		
NCI risk *			Normal	37	51
Standard	50	68	Enlarged <3cm	32	44
High	23	32	Enlarged >3cm	4	5
Specimens			Mediastinal mass		
Diagnosis	60	82	Not known	12	16
Relapse	13	18	No mass	58	80
Clinical outcome			Anterior mass	3	4
(diagnosis specimens)	(n=60)				
CCR †	11	18			
Relapsed	49	82			
Cytogenetics #					
<i>MLL</i>	1	1			
<i>BCR-ABL1</i>	7	10			
<i>TEL-AML1</i>	15	21			
<i>TCF3-PBX1</i>	2	3			
Hyperdiploid	26	36			
Other	22	30			

§ WBC: Peripheral white blood cell count at diagnosis, † CCR: Complete clinical remission, * NCI risk stratification: see Smith et al., (1996), # Cytogenetics: determined by karyotype or inferred by gene expression profiling, see Hoffman et al., (2006), ‡ BU/AU: below/above umbilicus.

Table 3.2 Characteristics of pre-B ALL specimens used for validation of *CTGF* expression

The panel of primary specimens used in this study consisted of 17 pre-B ALL, both diagnosis (D) and relapse (R) specimens, and 3 T-ALL specimens. *CTGF* expression measured by Affymetrix (U133A) human genome array and classed as high, med or low/negative based on probeset intensity (defined in Section 2.2.1.1).

Specimen Code	Diagnosis /Relapse	Lineage	cytogenetics (inferred) #	NCI Risk *	Age ‡ (months)	<i>CTGF</i> expression
BHDN9	D	B	<i>MLL</i>	High	166	high
[±] BHDR25	D	B	<i>BCR-ABL</i>	High	196	high
[±] BHRR25	R	B	<i>BCR-ABL</i>	High	201	high
BHDN12	D	B	HYP	High	29	high
BSDR85	D	B	<i>TEL-AML</i>	Standard	92	high
BSDN36	D	B	<i>TEL-AML</i>	Standard	64	high
BSDR79	D	B	<i>BCR-ABL</i>	Standard	113	med
BHRR83	R	B	HYP	High	250	med
BSDN34	D	B	<i>BCR-ABL</i>	Standard	50	med
BSRR89	R	B	<i>TCF3-PBX1</i>	Standard	133	med
BSDN67	D	B	Other	Standard	79	low/neg
BSDN40	D	B	Other	Standard	97	low/neg
BHDN7	D	B	Other	High	132	low/neg
BSDN51	D	B	Other	Standard	74	low/neg
BHDR31	D	B	Other	High	171	low/neg
BHRR92	R	B	Other	High	177	low/neg
BSDN63	D	B	Other	Standard	30	low/neg
THDN2	D	T	Other	High	70	low/neg
THDN4	D	T	Other	High	59	low/neg
THDN6	D	T	Other	High	161	low/neg

[±] Paired diagnosis and relapse specimens, # Cytogenetics inferred by 20 gene classifier method (Hoffmann et al. 2006), * NCI risk stratification, see Smith et al., (1996), ‡ Age in months at the time of diagnosis or relapse.

3.2.2 Validation of *CTGF* expression data

CTGF expression values generated by microarray (probeset 209101_at) and qRT-PCR were examined for concordance in these 20 ALL specimens. Figure 3.1 demonstrates *CTGF* gene expression levels and the correlation between the methods used to characterise expression in these samples. A comparison of *CTGF* expression values generated by microarray versus those obtained by qRT-PCR performed on freshly extracted, matched cryopreserved specimens, demonstrated a highly similar pattern of *CTGF* expression. Specimens BSDN36 and BSDR79 seemed to display different *CTGF* expression levels when measured by qRT-PCR, however overall a Pearson correlation of 0.972 was observed between methods, which was highly significant ($p < 0.001$). Those primary specimens that lacked one of the common cytogenetic rearrangements (Table 3.2 “Other”), were considered to be *CTGF*^{low/neg} when assessed by microarray, however, almost all of these cells exhibited an extremely low level of *CTGF* mRNA detectable by qRT-PCR, indicating the *CTGF* locus is in a permissive state.

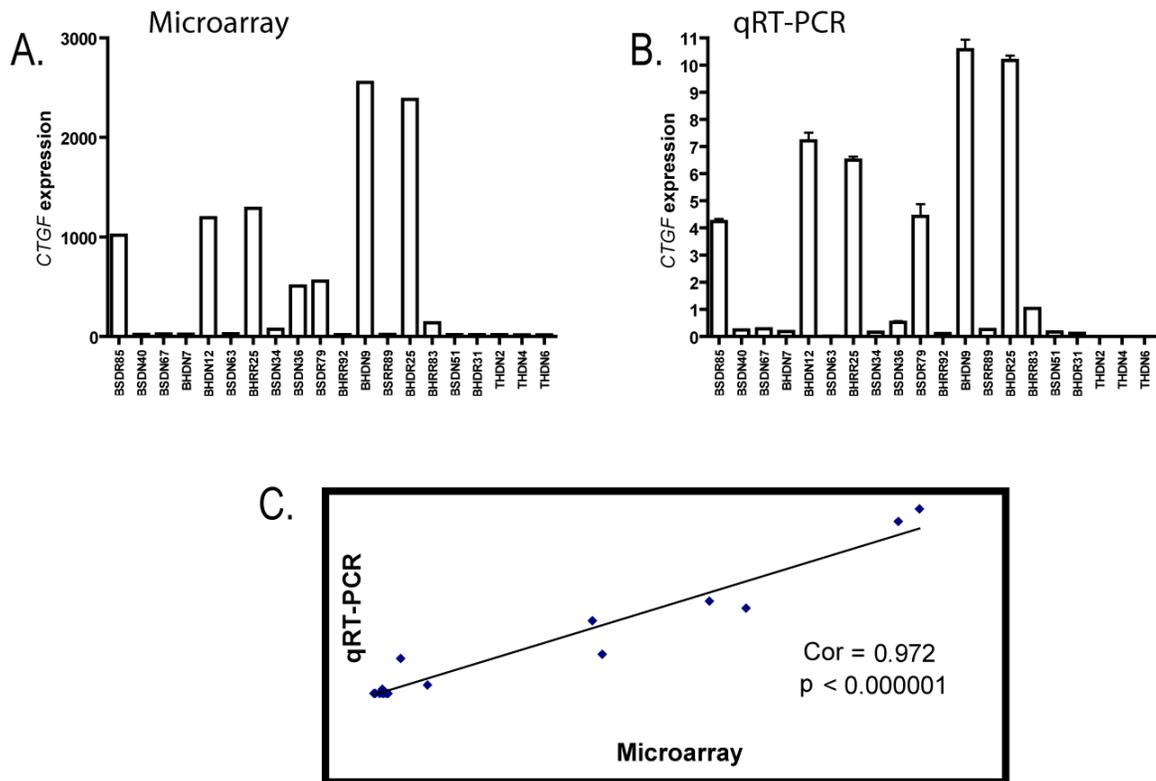


Figure 3.1 Analysis of *CTGF* expression by microarray and qRT-PCR

A. *CTGF* expression in a selection of primary B-lineage (n=17) and T-lineage (n=3) ALL specimens measured by Affymetrix human genome array U133A (probe 209101_at) **B.** *CTGF* expression measured in matched, cryopreserved specimens by qRT-PCR normalised to *ACTB*. Error bars represent the standard error of the mean of three technical replicates. **C.** Comparison of *CTGF* expression data generated by microarray and qRT-PCR. A highly significant Pearson correlation of 0.972 was observed (p<0.001) between the two methods.

3.2.3 Gene expression profiles associated with *CTGF* expression

Identifying genes that are co-regulated with *CTGF* may point to common promoter elements or trans-acting factors involved in activation of the *CTGF* promoter (Mansson et al. 2004). Probesets that correlated with *CTGF* expression in pre-B ALL were identified by examining global gene expression data from the PMH cohort (n=73), and two publically available paediatric pre-B ALL cohorts. These two cohorts were denoted

as the Ross cohort and the Kang cohort (Ross et al. 2003; Kang et al. 2010). The former was studied by Ross and colleagues while investigating the suitability of gene expression profiling for classifying paediatric ALL specimens into clinically relevant subtypes (Ross et al. 2003). This cohort consisted of 118 pre-B ALL specimens, *BCR-ABL1*; n=15, *TCF3-PBX1*; n=18, Hyperdiploid; n=15, *MLL* disease; n=20, *TEL-AML1*; n=20, and 28 cases where the aforementioned cytogenetic subtypes were not detected (Ross et al. 2003). The Kang cohort was comprised of 207 high-risk paediatric pre-B ALL patients enrolled in the Children's Oncology Group (COG) P9906 clinical trial and defined by age > 10yrs, WBC > $5 \times 10^4/\mu\text{l}$, and the absence of low-risk (Hyperdiploidy, *ETV6-RUNX1*) or very high-risk (hypodiploid, *BCR-ABL1*) disease (Kang et al. 2010). However, patients with low-risk cytogenetics were included if they displayed high-risk features like CNS or testicular involvement. While the Kang cohort was comprised of a discreet, high-risk group, it was considered appropriate for investigating *CTGF* correlates because *CTGF* was shown to be highly predictive of outcome in their study (Kang et al. 2010).

The distribution of *CTGF* gene expression values was assessed across samples in each cohort without normalizing the three datasets together, as the Kang cohort had been analysed using a different Affymetrix microarray platform (U133 plus2.0 array) to the PMH and Ross cohorts (U133A array). The range of *CTGF* expression in each cohort is shown in Figure 3.2A. A Spearman correlation analysis was performed on each cohort individually to examine which probesets displayed a similar pattern of expression to *CTGF*. The top 100 ranked probesets from the Spearman correlations were compared among cohorts and 11 probe sets overlapped between cohorts, as illustrated in Figure 3.2B. These probesets corresponded to 6 genes; suppressor of cytokine signalling 2 (*SOCS2*), myocyte enhancer factor 2C (*MEF2C*), adducin 3 (*ADD3*), gelsolin (*GSN*), metastasis suppressor 1 (*MTSS1*) and dihydropyrimidinase-like 2 (*DPYSL2*). These probesets displayed significant Spearman correlation values in the PMH cohort ranging from 0.68 ($p < 10^{-7}$) to 0.51 ($p < 10^{-5}$) and are shown in Figure 3.2C. The variance in the expression of these 11 probesets was compared to confirm that they display a similar pattern as *CTGF* (Figure 3.2D). All but one of the probesets, corresponding to *SOCS2*,

MEF2C, *ADD3*, *GSN* and *DPYSL2* appeared to have a distribution very similar to *CTGF* suggesting they may be co-regulated. In contrast, *MTSS1* gene expression varied substantially from *CTGF* in that it was expressed at comparatively lower levels and displayed less variance in expression among samples. As such this gene was omitted from further analysis.

Based upon the significant correlation with *CTGF* expression displayed by *SOCS2*, *MEF2C*, *ADD3*, *GSN* and *DPYSL2*, it was considered that there may be common *cis* regulatory elements in the 5' flanking region of these genes. One of these co-regulated genes, *MEF2C*, encodes a transcription factor, suggesting that it may itself be a candidate. However, analysis of the promoter regions of the above genes using the TFSEARCH database of transcription factor motifs available at <http://www.cbrc.jp/research/db/TFSEARCH.html>, did not identify any predicted *MEF2C* binding sites, even when the region examined was extended to include up to 4kb of 5' flanking sequence adjacent to the transcription start site.

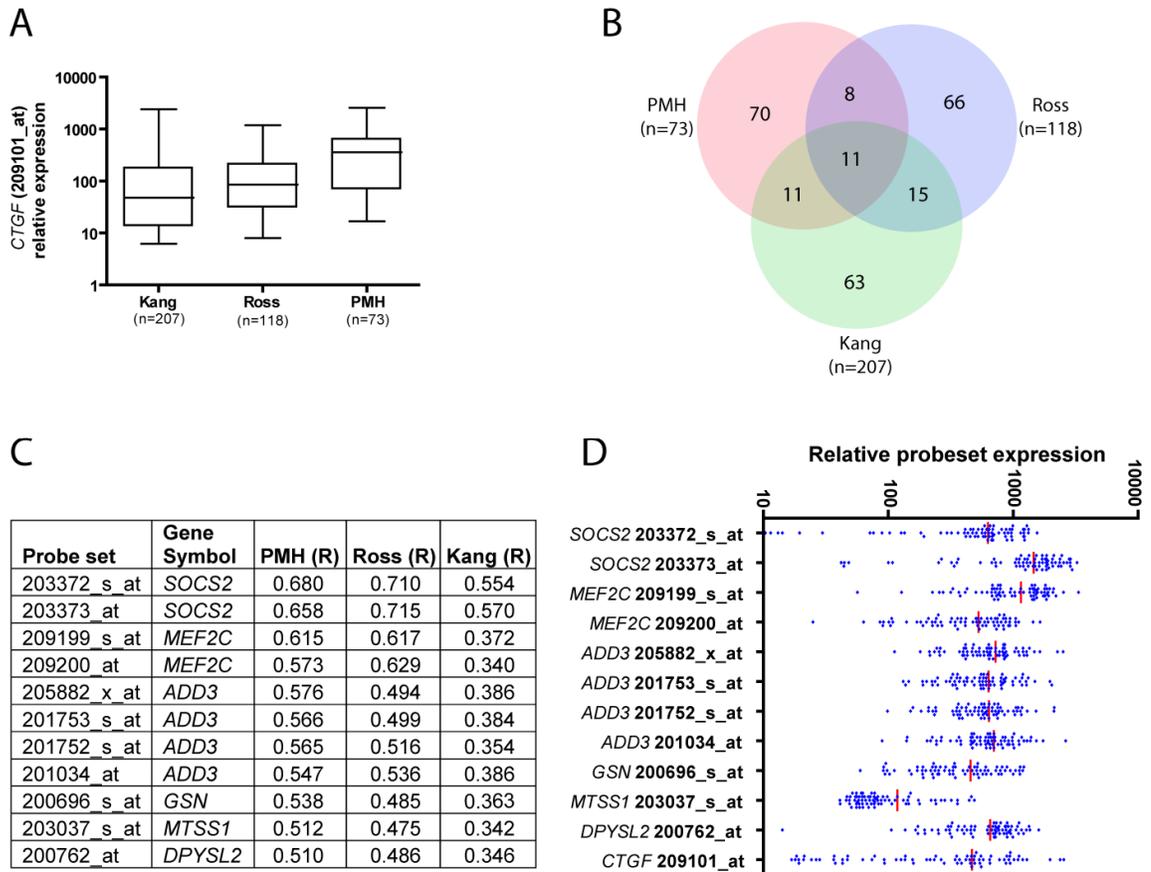


Figure 3.2 CTGF correlation analysis in three independent cohorts

A. Distribution of *CTGF* expression (probe set 209101_at) measured by microarray in three independent paediatric pre-B ALL cohorts. Cohort datasets are not normalised together as they originate from different Affymetrix platforms. **B.** A Spearman correlation analysis with *CTGF* (209101_at) was performed in each cohort individually and the top 100 probe sets from each cohort were compared. Eleven common probe sets were identified in the top 100 probe sets in each cohort. **C.** Overlapping probe sets corresponded to 6 genes; suppressor of cytokine signalling 2 (*SOCS2*), myocyte enhancer factor 2C (*MEF2C*), adducin 3 (*ADD3*), gelsolin (*GSN*), metastasis suppressor 1 (*MTSS1*) and dihydropyrimidinase-like 2 (*DPYSL2*). For each probe set the correlation with *CTGF* expression (R) is shown for each cohort. **D.** Expression of the 11 overlapping top-ranked probe sets as well as *CTGF* in the PMH cohort measured by microarray.

3.2.2.1 Common 5' flanking motifs in *CTGF*-correlated genes

In order to identify common sequence elements, the 5' flanking sequence of *CTGF* together with five highly correlated genes from -2 kb to +1 (relative to the transcriptional start site) were examined for shared sequence motifs using the MEME software suite (Bailey & Elkan 1994) available at <http://meme.nbcr.net/meme/intro.html>. This program searches for common sequence motifs in specified DNA sequences. Analysis with MEME identified two sequence motifs that were present in the 5' flanking sequence of the majority of the genes determined to be co-regulated with *CTGF*. These two motifs were referred to as MEME enriched motif 1 (MEM1) and MEM2.

The first motif, MEM1, was 15bp in length, heavily G-rich and was present in the promoters of *ADD3*, *DPYSL2*, *GSN*, *CTGF* and *SOCS2* (Figure 3.3A). This motif was examined for transcription factor binding sites using the TOMTOM motif comparison tool (Gupta et al. 2007) available as part of the MEME suite of applications. This program compares sequence motifs against known transcription factor binding matrices annotated in the TRANSFAC (Fu & Weng 2004) and JASPER databases (Portales-Casamar et al. 2010). The results of this analysis is shown in Figure 3.3B and predicted a binding site for the Ikaros family of transcription factors in MEM1 which are known to play an important role in haemopoietic development (Koipally et al. 1999). Although the Ikaros-binding site was predicted by the TOMTOM analysis, the Ikaros-binding motif was only a partial match to the 5' region of MEM1. Furthermore, the location of the MEM1 motif in the 5' flanking sequences shown in Figure 3.3 was highly variable, suggesting that this motif may not be a functional element, but rather a simple sequence repeat identified by the MEME analysis.

The second motif, MEM2, was 32 bp in length, as shown in Figure 3.4 and was identified at similar positions (approximately -650 to -1150) in the promoter regions of *ADD3*, *SOCS2*, *GSN* and *DPYSL2*, but was not present in the *CTGF* promoter. Significantly, this motif was also predicted to facilitate binding of the Ikaros family of

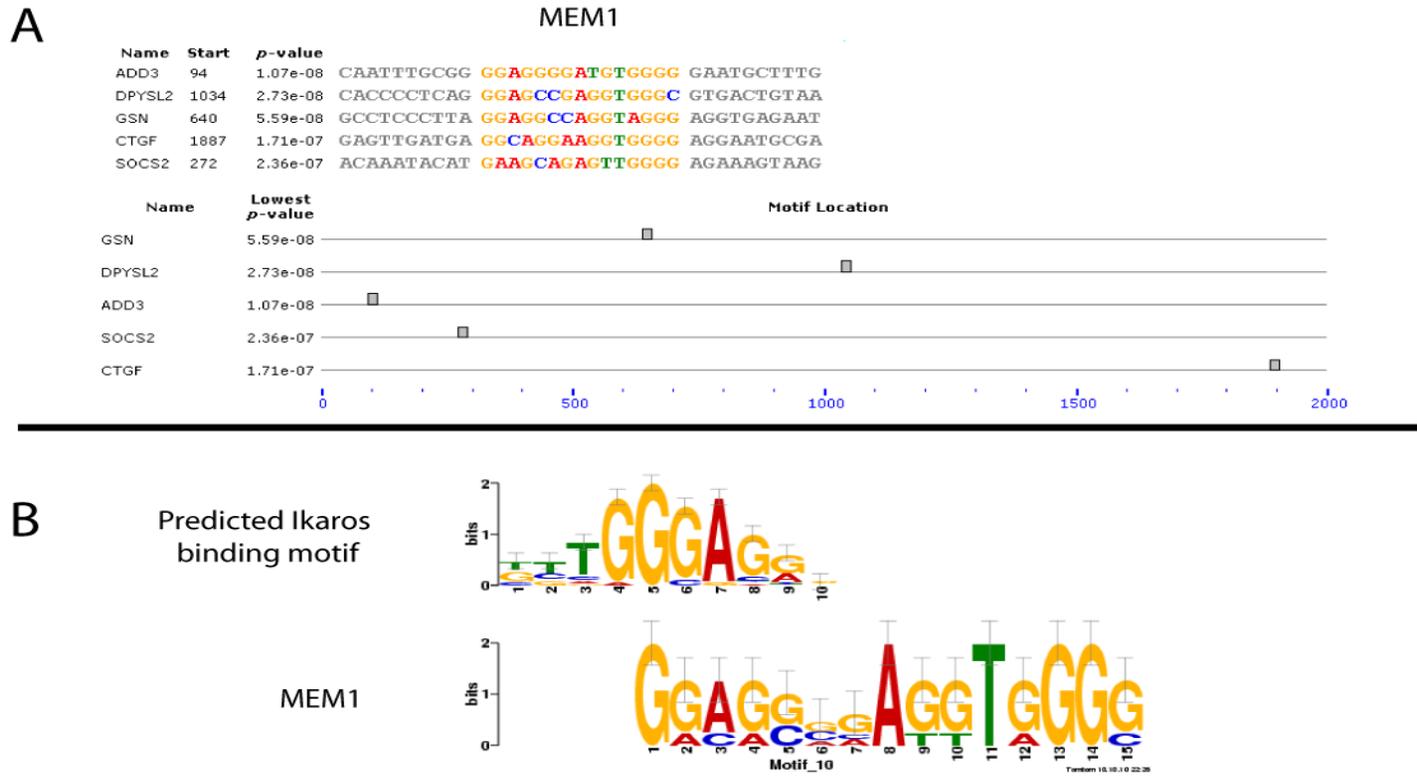


Figure 3.3 MEME enriched motif 1 (MEM1)

A. Top; Meme enriched motif (MEM) 1 and flanking sequence for each promoter including gene symbol (name), nucleotide coordinate (start) of motif in input sequence (-2kb to +1) and p-value of motif discovery. Sequences are ranked from top to bottom by statistical significance. Bottom; location of motif in input sequence. **B.** Top; Predicted Ikaros-binding motif. Bottom; MEM1 and ideogram representing nucleotide incidence at each base in motif. Analysis was performed using MEME and TOMTOM programs available online at <http://meme.nbcr.net/meme/intro.html>.

A

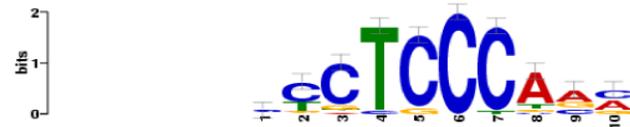
MEM2

Name	Start	p-value	Sequence
ADD3	1341	8.25e-30	GATCTACCTG CCTCAGCCTCCCAAAGTGCTGGGATTACAGGCCTGAGCCACCACGCCCGG CAATATTGTG
SOCS2	843	3.40e-28	GATCCACCCG CCTCAACCTCCCAAAGTGCTGGGATTACAGGCCTGAGCCACCACGCCCGG CCAGGATTCT
GSN	1060	5.03e-20	CCGGAGTCAC CCGGCAACTCCAAAAGTCTGGGAACTCATGTGTGCACCAGGACTGCTGG GAGGCCGTCT
DPYSL2	941	4.04e-19	CACTGTAAC CCTCAGGACCCCGGGCGCTGGGATCGCAGAGCTGCCCCGACAGCGCT GCAGGCACCA



B

Predicted Ikaros binding motif



MEM2



Figure 3.4 MEME enriched motif 2 (MEM2)

A. Top; Meme enriched motif (MEM) 2 and flanking sequence for each promoter including gene symbol (name), nucleotide coordinate (start) of motif in input sequence (-2kb to +1) and p-value of motif discovery. Sequences are ranked from top to bottom by statistical significance. Bottom; location of motif in input sequence. **B.** Top; Predicted Ikaros-binding motif. Bottom; MEM2 and ideogram representing nucleotide incidence at each base in motif. Analysis was performed using MEME and TOMTOM programs available online at <http://meme.nbcr.net/meme/intro.html>.

transcription factors and displayed higher homology with the IKAROS motif than MEM1. This motif additionally contained an invariant CTGGGA sequence in all four promoters, however this sequence element was not predicted to be part of any transcription factor binding motif by TOMTOM. This was surprising, as the first three members of the Ikaros family; IKAROS, HELIOS and AIOLOS, all contain a central GGGGA sequence in their DNA recognition sequences (shown in Figure 3.5) and the invariant CTGGGA identified in MEM2 is compatible with IKAROS, HELIOS and AIOLOS binding matrices (Molnar & Georgopoulos 1994). While the data provided by analysis with MEME and TOMTOM were novel, the existence of Ikaros family binding sites in these genes warranted further validation.

IKZF1 - IKAROS						IKZF2 - HELIOS						IKZF3 - AIOLOS					
pos	A	C	G	T	consensus	pos	A	C	G	T	consensus	pos	A	C	G	T	consensus
1	7	5	7	5	N	1	9	10	12	5	N	1	3	0	7	15	T
2	6	9	0	9	N	2	11	5	7	13	N	2	8	8	5	4	N
3	3	3	7	11	N	3	5	7	7	17	N	3	0	13	3	9	Y
4	2	6	1	15	T	4	6	10	0	20	Y	4	6	5	0	14	T
5	2	0	22	0	G	5	6	0	30	0	G	5	5	0	20	0	G
6	0	0	24	0	G	6	0	0	36	0	G	6	0	0	25	0	G
7	0	0	24	0	G	7	0	0	36	0	G	7	0	0	25	0	G
8	24	0	0	0	A	8	36	0	0	0	A	8	25	0	0	0	A
9	22	0	0	2	A	9	18	5	1	12	W	9	25	0	0	0	A
10	3	4	2	15	T	10	10	7	6	13	N	10	0	4	0	18	T
11	11	2	8	3	R	11	12	8	9	7	N	11	16	0	5	4	A
12	3	13	2	6	C	12	9	18	5	4	N	12	1	18	0	6	C
13	3	14	3	4	C							13	1	17	0	7	C

Figure 3.5 DNA binding matrices for the Ikaros family members 1, 2 and 3

DNA binding matrices are shown for the first three members of the Ikaros family of transcription factors. These members are IKAROS (IKZF1), HELIOS (IKZF2) and AIOLOS (IKZF3). Figure adapted from Molnar et al., (1994).

3.2.2.2 Analysis of putative Ikaros family binding sites

The *In silico* evidence presented above suggested that members of the Ikaros family of transcription factors IKAROS, HELIOS and AIOLOS may bind to the 5' flanking sequence of genes displaying similar patterns of expression as *CTGF*. Each promoter was independently examined for predicted Ikaros family binding sites using the TFSEARCH transcription factor motif database (<http://www.cbrc.jp/research/db/TFSEARCH.html>). The 5' flanking sequence of *GSN*,

DPYSL2, *ADD3*, *SOCS2*, *MEF2C* and *CTGF* were all predicted to contain at least one, and in some cases several, binding sites for the Ikaros family member HELIOS as outlined in Figure 3.6. A number of sites co-localised with the matrices predicted by MEME in Figures 3.3 and 3.4. These sites, annotated in Figure 3.6 included 1, 2, 3, 4, 8, 10 and 12. These predicted binding sites are based on functionally characterised murine Helios binding motifs (Molnar & Georgopoulos 1994), however the sequence identity shared between the human and murine *HELIOS/Helios* coding region is over 97% and they are predicted to recognise similar, if not identical, recognition sequences (Hosokawa et al. 1999). Furthermore, some of the predicted binding sites were also compatible with binding by IKAROS. The remarkable similarities in the binding matrices for IKAROS, HELIOS and AIOLOS as outlined in Figure 3.5 suggested that while some predicted binding sites are only compatible with HELIOS or IKAROS, any of these three proteins may in fact bind to the sites identified in the promoters of genes that showed expression correlated to *CTGF* gene expression.

3.2.2.3 Expression of Ikaros family genes in the PMH cohort

Binding motifs for HELIOS and IKAROS were identified in the 5' flanking sequence of genes that correlate with *CTGF* expression, suggesting a role for the Ikaros family of transcription factors in regulating expression of these genes. The PMH microarray dataset was examined to determine whether changes in expression of any Ikaros family members; *IKAROS*, *HELIOS*, *AIOLOS*, *EOS* and *PEGASUS* was associated with *CTGF* expression in this cohort (Figure 3.7). Specimens were grouped into *CTGF*^{pos} and *CTGF*^{low/neg} groups based on microarray expression values, and mean expression of the Ikaros family of genes were compared between these groups. All Ikaros family genes were expressed except for *AIOLOS* which displayed probeset intensity values well below what would normally be classed as a background signal on the U133A chip (intensity value of <25). There was no significant difference observed in the mean expression of any Ikaros family member between *CTGF*^{pos} and *CTGF*^{low/neg} specimens as shown in Figure 3.7, suggesting that any effect mediated by HELIOS or other Ikaros proteins was not due to alterations in their mRNA expression levels.

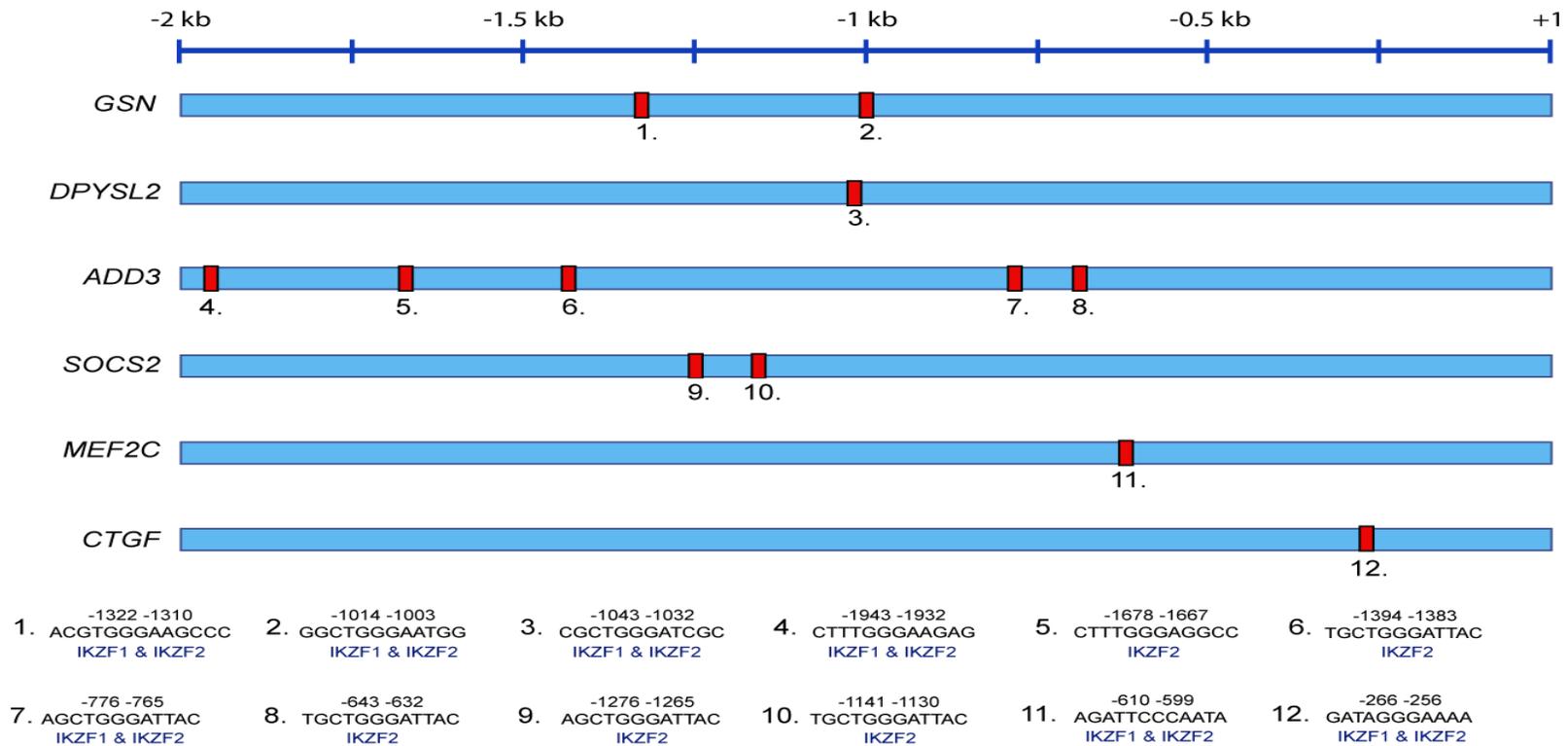


Figure 3.6 Helios (IKZF2) is predicted to bind the promoters of *CTGF* and correlated genes

Binding sites for IKZF1/IKAROS and IKZF2/HELIOS predicted in the 5' flanking regions (-2 kb to +1) of *CTGF*-correlated genes by TFSEARCH. Numbered red boxes indicate predicted binding sites (1-12) which are annotated (bottom) with coordinates relative to the transcription start site of each gene, motif sequence, and indication of which factor was predicted to bind the target.

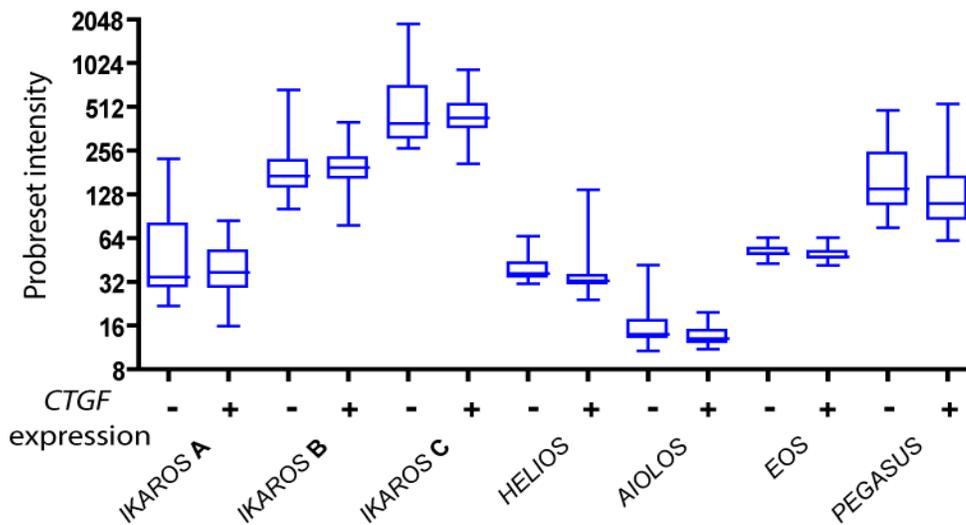


Figure 3.7 Expression of Ikaros family members in the PMH cohort

Expression of Ikaros transcription factors, measured by microarray analysis of primary specimens in the PMH cohort. Specimens were grouped according to *CTGF* expression status (samples classed as *CTGF*^{low/neg} are denoted by the minus symbol). Expression indicated by probeset intensity for *IKAROS* probesets; A-216901_s_at, B-205038_at and C-205039_s_at, as well as *HELIOS* 220567_at, *AIOLOS* 221092_at, *EOS* 208472_at and *PEGASUS* 220086_at.

3.2.2.4 *IKAROS* gene lesions are associated with *CTGF* expression

Deletions and mutations of *IKAROS* are common events in high-risk pre-B ALL (Kuiper et al. 2010; Mullighan et al. 2008), and internal in-frame deletions can produce dominant-negative (DN) protein isoforms that can inhibit the function of Ikaros proteins, in the absence of changes in gene expression (Sun et al. 1996). The Ikaros family of proteins is thought to promote gene silencing through enforcing repressive chromatin marks at target gene loci (Koipally et al. 1999). If aberrant *CTGF* expression can indeed be attributed to a loss of function in Ikaros family genes in pre-B ALL, then those specimens harbouring mutations or deletion of *IKAROS* would be expected to exhibit high *CTGF* expression. A recent report from St Jude Children’s Research Hospital

examined the prognostic significance of *IKAROS* deletions and mutations, herein referred to collectively as *IKAROS* alterations, using a combination of SNP arrays and direct sequencing of B-cell developmental genes (Mullighan et al. 2009). This study was undertaken in a selection of high-risk patients from the COG P9906 clinical trial cohort. Patient specimens used for the study by Kang and colleagues were also selected from the COG P9906 cohort (Kang et al. 2010). Despite the selection criteria of the two studies being very similar, focussing on high-risk pre-B ALL, not all specimens examined for copy number alterations by Mullighan *et al* were represented in the gene expression data associated with the Kang cohort. However, of the 221 specimens in the Mullighan cohort, 198 matched with gene expression data from the Kang cohort. Pairing of this data allowed the exploration of the effect of *IKAROS* alterations upon *CTGF* expression in high-risk pre-B ALL.

Of the 67 patient specimens that were identified as harbouring an *IKAROS* alteration by Mullighan *et al*, gene expression data from the Kang cohort was available for 59 of these. To examine whether *IKAROS* alterations were associated with aberrant *CTGF* expression, mean *CTGF* expression was compared between these 59 specimens and the remainder of specimens in the Mullighan cohort for which gene expression data was available. This analysis shown in Figure 3.8A demonstrated that *IKAROS* alterations were indeed associated with higher *CTGF* expression, with a significant difference in mean *CTGF* expression evident between these two groups ($p=0.036$).

CTGF expression was also investigated in a more homogeneous subset of pre-B ALLs, which all contained a deletion in at least one critical B-cell developmental gene specimens (66.5 % of high-risk cases), including *IKAROS*, *PAX5*, *TCF3*, *EBF1*, *RAG1/2*, *BLNK*, *BCL11A*, *HELIOS*, *LEF1*, *MEF2C*, *SOX4* and *SPI1* (Mullighan et al. 2009). *IKAROS* deletions were present in 43% of this group compared to 29% in the total P9906 cohort, thus it was expected that any association between *IKAROS* alterations and *CTGF* gene expression would be easily observed in this group. These specimens were stratified based upon the presence or absence of *IKAROS* alterations and examined for *CTGF* expression, shown in Figure 3.8B. A total of 154 specimens

harboured disruption of one or more of these critical B-cell developmental genes. Of these 154 specimens, gene expression data was available for 137 specimens, and 59 of these were affected by *IKAROS* alterations. This group of 59 specimens exhibited higher mean *CTGF* expression compared to those specimens harbouring B-cell pathway lesions but no *IKAROS* alterations and this was highly significant ($p < 0.001$). These findings support the hypothesis that *IKAROS* alterations are associated with aberrant *CTGF* expression in high-risk pre-B ALL.

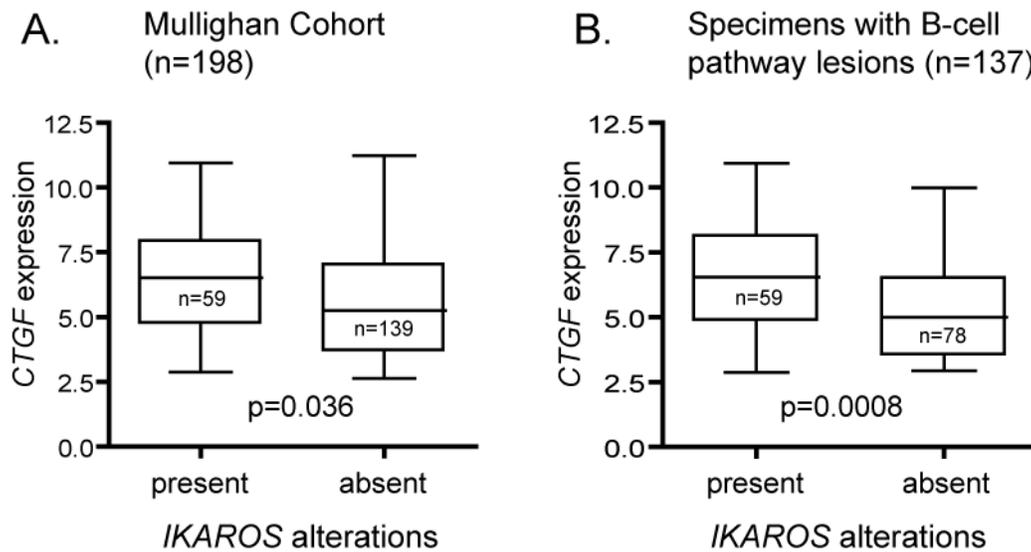


Figure 3.8 *IKAROS* alterations are associated with aberrant *CTGF* expression in high-risk pre-B ALL

A. Patient specimens from the Mullighan cohort (Mullighan et al. 2009), were divided into two groups based on whether they had a confirmed *IKAROS* alteration, then examined for expression of *CTGF*. Those patients with alterations affecting *IKAROS* (n=59) displayed significantly higher *CTGF* expression ($p=0.036$) compared to the remainder of the cohort in which alterations of the *IKAROS* locus were absent. **B.** *CTGF* expression was examined in patient specimens harbouring disruption of at least one critical B-cell pathway gene, including *IKAROS*, *PAX5*, *TCF3*, *EBF1*, *RAG1/2*, *BLNK*, *BCL11A*, *HELIOS*, *LEF1*, *MEF2C*, *SOX4* and *SPI1*. Of the 137 specimens harbouring mutation of at least one of these genes, 59 had a deletion or mutation affecting the

IKAROS gene, and these specimens had significantly higher *CTGF* expression ($p < 0.001$).

3.2.3 Clinical features associated with *CTGF* expression

CTGF expression reportedly has prognostic significance in pre-B ALL and has been associated with reduced RFS for some forms of the disease in both adults and children (Sala-Torra et al. 2007; Kang et al. 2010). As shown in Figure 3.9, *CTGF* expression in the PMH cohort varied considerably across specimens. This suggested that *CTGF* expression might correlate with one or more clinical features present at the time of diagnosis, thus acting as a biomarker for a specific group (defined by age, cytogenetic group, etc) at greater risk (Yoshida et al. 2009; Choi et al. 2006). Patients comprising the PMH cohort were diagnosed locally, enabling access to detailed patient data for this study.

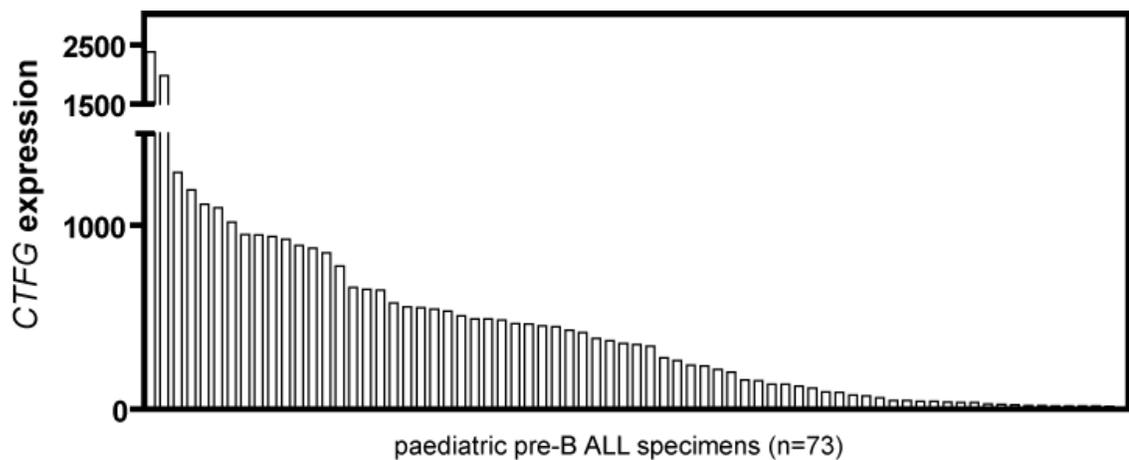


Figure 3.9 *CTGF* gene expression in the PMH cohort

CTGF gene expression measured by Affymetrix human genome array U133A (probeset 209101_at) in the paediatric pre-B ALL cohort from PMH (n=73).

Patient specimens obtained at the time of diagnosis (n=60) were stratified based upon clinical features including gender (male; n=33, female; n=27), age at diagnosis in years (1.5-3 yrs; n=12, 3-6 yrs; n=16, 6-10 yrs; n=22 and 10-20 yrs; n=10), percentage blasts in bone marrow (>95%; n=38, 90-95%; n=12 and <90%; n=10), haemoglobin levels in g/dL, (>10; n=12, 7.5-10; n=18, 7.5-5; n=25 and <5; n=5), lymph node status (enlarged; n=33, normal; n=27) and spleen status (enlarged; n=38, normal; n=22). *CTGF* expression was also compared between these 60 diagnosis specimens and 13 relapse specimens. The mean *CTGF* expression was compared between these clinically relevant groups as shown in Figure 3.10. Patients with enlarged lymph nodes displayed significantly lower *CTGF* expression (p=0.0349) than those that did not have lymph node involvement. A similar trend was observed for those patients with an enlarged spleen, however this did not reach significance. No other clinical features examined demonstrated an association with *CTGF* expression.

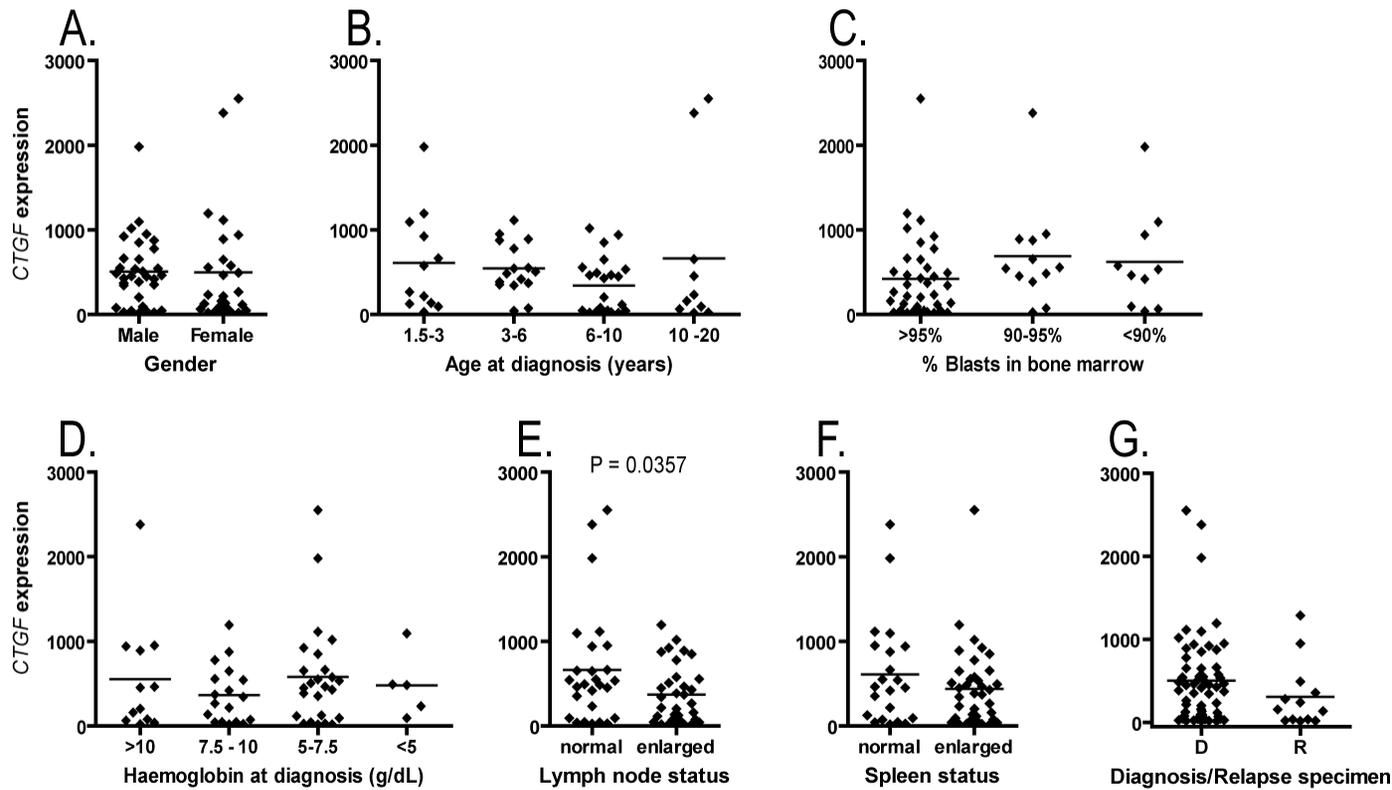


Figure 3.10 Comparison of *CTGF* expression versus clinical features present at diagnosis in the PMH cohort

CTGF gene expression (measured by microarray) in pre-B ALL specimens (n=60) stratified by clinical features present in patients at the time of diagnosis, including gender (A.), age at diagnosis (B.), percentage blasts in bone marrow (C.), haemoglobin (D.), lymph node status (E.) and Spleen status (F.). *CTGF* expression was also compared between the 60 diagnosis specimens and 13 relapse specimens (G.). Mean *CTGF* expression was compared between groups using unpaired t-tests or one-way ANOVA where appropriate.

3.2.4 Prognostic significance of *CTGF* expression

The association between *CTGF* expression in specimens obtained at diagnosis in the PMH cohort and both patient relapse and overall survival was examined to investigate the prognostic relevance of *CTGF* expression (Figure 3.11). Patient diagnostic specimens were stratified based upon *CTGF* expression into two groups, *CTGF*^{pos} and *CTGF*^{low/neg}, based on previously determined gene expression cut-offs (section 2.2.1.1), and the incidence of relapse and overall survival was examined. Three patients were censored from the analysis of overall survival due to incomplete survival data for the 5yr period. There was a trend toward lower RFS in the *CTGF*^{pos} group at 71% compared to 83% for the *CTGF*^{low/neg} group, however this was not statistically significant (p=0.39). The overall 5 year survival did not differ between the *CTGF*^{pos} and *CTGF*^{low/neg} groups. The trend for higher relapse lends support to previous studies indicating that *CTGF* has prognostic significance in pre-B ALL (Sala-Torra et al. 2007; Kang et al. 2010).

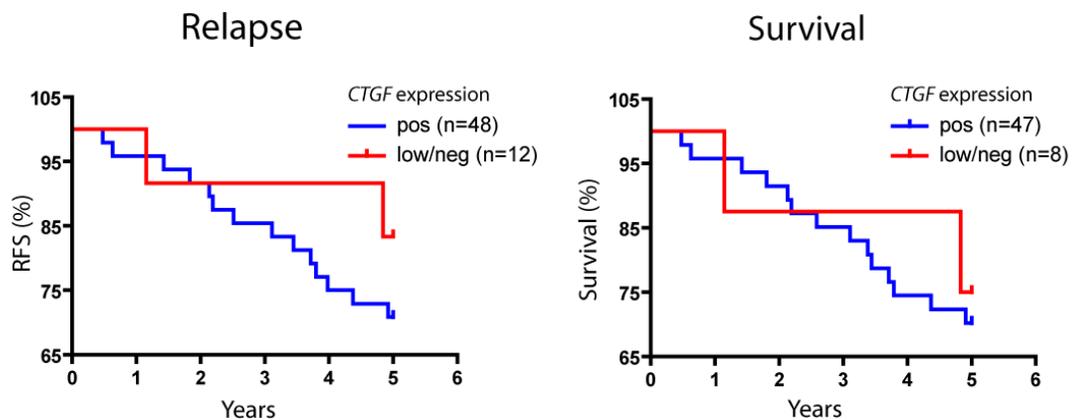


Figure 3.11 Association between *CTGF* expression and patient outcome.

Patients were stratified based upon *CTGF* expression (pos or low/neg) and assessed for both relapse-free survival (RFS) and overall survival over 5 years.

3.3 Discussion

Deregulated *CTGF* expression in pre-B ALL has now been identified by several studies, however the mechanisms driving its aberrant expression remain unknown (Vorwerk et al. 2000; Sala-Torra et al. 2007; Boag et al. 2007). Before analysing global gene expression data and investigating clinical correlates in the PMH cohort, *CTGF* expression levels obtained by microarray were validated in cryopreserved pre-B ALL specimens by qRT-PCR. These validation experiments revealed that there was a significant correlation between microarray and qRT-PCR methods for detection of *CTGF* mRNA, confirming that the global gene expression data obtained by Boag and colleagues provided an accurate representation of *CTGF* gene expression in the PMH cohort (Boag et al. 2007).

Analysis of *CTGF* mRNA levels by qRT-PCR revealed that those primary samples that were not grouped into one of the common cytogenetic groups exhibited very low, almost undetectable *CTGF* expression. These specimens had previously been classified as *CTGF*^{low/neg} by array, based on *CTGF* probeset signals that were close to or the same as background levels. However, the extreme sensitivity of qRT-PCR confirmed that there was low level *CTGF* expression in almost all of these primary specimens. The level of *CTGF* mRNA that was detected was so low as to be unlikely to result in translation of CTGF protein, however this finding nonetheless suggests that the *CTGF* locus was in a permissive transcriptional state in these pre-B ALL specimens. These data infer that biological changes associated with *BCR-ABL1*, *TEL-AML1*, *MLL* disease and hyperdiploidy may be important in actively promoting overexpression of *CTGF*, although the underlying factors involved remain undefined. Recurrent oncogenic mutations and genomic rearrangements are common in pre-B ALL (Mullighan et al. 2007; Pui 2006), and if such genomic lesions affect the *CTGF* locus at 6q23.1 they may enhance *CTGF* gene expression. Extensive genome-wide analysis of genetic alterations in pre-B ALL have thus far failed to identify any recurrent alterations at the *CTGF* locus (Mullighan et al. 2007; Kuiper et al. 2007b), however a more focussed investigation of this region is warranted.

3.3.1 Clinical correlates of *CTGF*

The vast range of *CTGF* expression observed in the PMH cohort prompted the investigation into associations between clinical features present at the time of diagnosis and *CTGF* gene expression in patient specimens. When patients were stratified based on clinical features there was a significant association between lymphadenopathy and lower *CTGF* mRNA expression in patient specimens ($p=0.036$). There was also a similar trend with splenomegaly, however this did not reach significance. High *CTGF* expression has been positively associated with lymph node metastasis in squamous cell carcinomas of the neck (Li et al. 2007), while low *CTGF* expression is associated with lymph node metastasis in colorectal cancer (Lin et al. 2005). Thus the relationship between *CTGF* expression and lymph node metastasis is controversial and is likely to depend heavily upon tumour type and cell of origin. Validation of the association between lymphadenopathy and low *CTGF* expression in pre-B ALL should be performed in an independent patient cohort, however no such patient data was available during this study.

Details from unpublished experiments performed in the laboratory of Prof Ursula Kees imply a role for CTGF in enhancing adhesion of pre-B ALL cells to stromal elements. This finding coupled with the high incidence of bone marrow fibrosis in ALL (Kundel et al. 1964; Wallis & Reid 1989) and correlation between bone marrow fibrosis and high MRD after induction therapy (Noren-Nystrom et al. 2008), suggests that fibrotic or reactive stroma in the bone marrow microenvironment may represent a preferred sanctuary site for leukaemic blasts compared to peripheral organs. It is conceivable that aberrant CTGF secretion within the bone marrow space may result in the formation of reactive stroma, thus promoting homing of leukaemic cells to the bone marrow, rather than peripheral organs. Future functional studies focussing on the effect of CTGF-induced alterations in bone marrow stroma are expected to test such hypotheses and may confirm a role for CTGF in attracting leukaemic cells to this microenvironment.

3.3.2 *CTGF* expression is a prognostic biomarker

The prognostic importance of *CTGF* expression in pre-B ALL has been demonstrated in both adult and paediatric pre-B ALL, with *CTGF* expression correlating with low RFS (Sala-Torra et al. 2007; Kang et al. 2010). Analysis of the role of *CTGF* as a biomarker in the PMH cohort confirmed an association between *CTGF* expression and a reduced RFS. While the difference in RFS between *CTGF*^{pos} and *CTGF*^{low/neg} groups did not reach statistical significance, the present findings support those made by Kang and colleagues that *CTGF* expression can be predictive of relapse in paediatric pre-B ALL (Kang et al. 2010). The striking associations observed in the Kang and Sala-Torra studies were not observed in the PMH cohort, and this may be due to differences in treatment protocols, as well as differences in the molecular types represented in the various cohorts.

While both the Kang and PMH cohorts were comprised of paediatric pre-B ALL patients, there were significant differences in the composition of these cohorts. The Kang cohort was comprised of high-risk patients as defined by age > 10yrs, WBC > 5x10⁴ /µl, and the absence of low-risk (Hyperdiploidy, *ETV6-RUNX1*) or very high-risk (hypodiploid, *BCR-ABL1*) disease (Kang et al. 2010), whereas the PMH cohort was comprised predominantly of standard-risk patients at a ratio of 3:1 (standard:high). Furthermore, the PMH cohort is comprised predominantly of younger patients with 73% aged between 1 and 10 years, and this age group has a favourable outcome compared to those over 10 years of age (Pieters & Carroll 2010). It is unsurprising therefore that the ability of *CTGF* mRNA to independently predict relapse was more limited in the PMH cohort, as the majority of patients in this cohort were predicted to have a favourable outcome based on the assessment of risk-features alone. Analysis of clinical features associated with *CTGF* expression would be useful in the Kang cohort which consists of a highly specific subtype of ALL, however this data was not made publically available (Kang et al. 2010). Analysis of such a discreet prognostic group may infer an association between *CTGF* expression and biological features and suggest a functional role for *CTGF* in pre-B ALL.

3.3.3 Gene expression correlates of *CTGF*

In order to investigate regulation of the *CTGF* locus in pre-B ALL, interrogation of gene expression profiles correlated with that of *CTGF* was undertaken in the PMH cohort (Boag et al. 2007). Two other publically available pre-B ALL microarray cohorts were also included to ensure robust analysis (Ross et al. 2003; Kang et al. 2010). Identification of genes highly correlated with *CTGF* expression was expected to identify common transcriptional networks (Mansson et al. 2004). The five genes that demonstrated significant correlation with *CTGF* expression in all three cohorts (*SOCS2*, *MEF2C*, *ADD3*, *GSN*, and *DPYSL2*) have disparate functions, and while one of these genes, *MEF2C* encodes a transcription factor, direct involvement of MEF2C protein was ruled out due to the lack of predicted MEF2C binding sites in the 5' flanking sequence of these *CTGF*-correlated genes.

Analysis of the 5' flanking sequence of *CTGF* and the five *CTGF*-correlated genes revealed predicted binding sites for the Ikaros family of transcription factors in all 6 sequences. At least one HELIOS binding site was identified in the 5' flanking regions of all six genes, and binding motifs for IKAROS in all but the *SOCS2* flanking sequence. These predicted sites were in agreement with published IKAROS and HELIOS binding matrices (Molnar & Georgopoulos 1994). The first four members of the Ikaros family, IKAROS, HELIOS, AIOLOS and EOS, have similar DNA binding preferences through recognition of a core GGGA[A/T] pentamer motif (Molnar & Georgopoulos 1994; Perdomo et al. 2000). The fifth member; PEGASUS has very different requirements, preferring an 8pb GNNGTGNG sequence motif (Perdomo et al. 2000).

3.3.4 Ikaros family of transcriptional regulators

As discussed in Chapter 1, Ikaros proteins are sequence specific transcription factors required for normal haemopoiesis and lymphoid specification, that can both repress and potentiate gene expression through interactions with DNA and chromatin modifying complexes (Koipally et al. 1999; Kim et al. 1999; Cobb & Smale 2005). Ikaros proteins function as homodimers, but can also form heterodimers with other Ikaros family

members (Rebollo & Schmitt 2003). Mutations or in-frame deletions affecting the N-terminal zinc finger DNA-binding domain can result in the synthesis of dominant negative (DN) isoforms that bind to canonical IKAROS (or other Ikaros family proteins) and inhibit the ability of the protein dimer to bind to DNA target sequences (Sun et al. 1996). Forced expression of DN IKAROS mutants immortalises murine haemopoietic precursor cells (Ruiz et al. 2008) and loss of function mutations affecting Ikaros proteins are associated with haematological malignancies including AML, pre-B ALL and T-cell leukaemia (Yagi et al. 2002; Mullighan et al. 2007; Tabayashi et al. 2007).

Recent reports have identified deletion of IKAROS or expression of DN isoforms as a powerful indicator of relapse and a poor prognosis in high-risk pre-B ALL (Kuiper et al. 2010; Kang et al. 2010; Mullighan et al. 2008; Mullighan et al. 2009). In the present study, the relative expression levels of Ikaros family genes remained constant between $CTGF^{\text{pos}}$ and $CTGF^{\text{low/neg}}$ specimens in the PMH cohort, thus gross changes in expression of Ikaros family genes cannot account for changes in $CTGF$ expression. Analysis of the Mullighan cohort of high-risk, paediatric, pre-B ALL patients (Mullighan et al. 2009), revealed that $CTGF$ expression was significantly higher in those high-risk patients that harboured deletion or mutation of the *IKAROS* gene, compared to those with wild-type *IKAROS*. Furthermore, analysis of a sub-set of specimens harbouring genetic lesions in critical B-cell developmental genes revealed a highly significant association between *IKAROS* alterations and high $CTGF$ expression.

These data provide independent evidence to support the hypothesis that Ikaros family proteins are linked to aberrant $CTGF$ gene expression. A more detailed analysis of the role of the Ikaros family of transcriptional regulators in the context of $CTGF$ expression is warranted. Furthermore, the focus of any future study should be extended to include the first three members of the Ikaros family of proteins (*IKAROS*, *AIOLOS* and *HELIOS*), because of the similarities in their DNA binding preferences (Molnar & Georgopoulos 1994; Rebollo & Schmitt 2003), their ability to interact with and modulate the function of each other (Sun et al. 1996), as well as the presence of predicted HELIOS binding motifs in the 5' flanking sequence of all $CTGF$ -correlated

genes investigated. Such a study may uncover as-yet unidentified DN forms of other Ikaros family proteins that are of importance in modulating IKAROS function.

3.3.6 Conclusions

In summary, the findings reported in this chapter suggest that the *CTGF* locus exists in a transcriptionally permissive state in the majority of pre-B ALL specimens. A panel of five genes displayed significant correlation with *CTGF* expression in three independent patient cohorts and these five genes, as well as *CTGF* all contain predicted binding sites for the Ikaros family transcription factors IKAROS and HELIOS in their 5' flanking region. *IKAROS* mutations are significantly associated with increased *CTGF* expression in high-risk pre-B ALL, particularly in those patients with lesions affecting B-cell developmental pathways. Analysis of RFS survival in the PMH cohort demonstrated a lower RFS in patients exhibiting high *CTGF* expression, confirming findings by others in adult and paediatric ALL (Sala-Torra et al. 2007; Kang et al. 2010).

Chapter 4

Sequence and Structure of the *CTGF* Locus in Pre-B ALL

Chapter 4

Sequence and Structure of the *CTGF* Locus in Pre-B ALL

4.1 Introduction

CTGF is overexpressed in a high proportion of pre-B ALL tumours (Boag et al. 2007; Sala-Torra et al. 2007), however the mechanisms promoting its aberrant expression are unclear. The answer to this question may lie in DNA mutations or genomic rearrangements affecting the *CTGF* locus. Discrete DNA mutations can alter the ability of nuclear factors to bind to DNA and regulate gene expression. Mutation within the *CTGF* promoter or 5' flanking sequence could result in enhanced expression by disrupting sequences normally bound by repressive transcription factors such as the putative Ikaros family binding site identified in the previous chapter (Laurila & Lahdesmaki 2009; Blesa et al. 2008). Alternatively, mutation may create a new transcription factor binding site, as seen in the mutational activation of the survivin (*BIRC5*) promoter in some breast cancers (Boidot et al. 2010) or aberrant activation of superoxide dismutase 2 (*SOD2*) promoter in a broad range of carcinomas (Xu et al. 2008). As well as mutations involving the 5' flanking sequence of *CTGF*, disruption of regulatory elements in the 3' UTR may also serve to enhance *CTGF* expression by disrupting post-transcriptional, miRNA-mediated repression. Altered patterns of gene expression have been attributed to changes in miRNA levels in numerous cancers including leukaemia (Tanaka et al. 2009; Agueli et al. 2010; Volinia et al. 2010). Mutation of miRNA binding sites may result in altered *CTGF* expression or mRNA stability in pre-B ALL cells.

A feature of the genomic landscape of leukaemia is the existence of recurrent genomic rearrangements that can influence gene expression or create oncogenic fusion proteins (Pui 2006). Whole genome scans can now be performed on large patient cohorts using high resolution single nucleotide polymorphism (SNP) arrays or comparative genome hybridisation (CGH) to identify regions frequently affected by copy number alterations (CNAs) in the cancer genome. These studies have recently highlighted genetic loci that

are frequently deleted in paediatric pre-B ALL including *PAX5*, *EBF* and *IKAROS* (Mullighan et al. 2007; Harvey et al. 2010; Kuiper et al. 2007b). The long arm of chromosome 6 is frequently altered in pre-B ALL, but while the most common alterations are deletions, these events rarely affect the *CTGF* locus at 6q23.1 (Thelander et al. 2008; Mullighan et al. 2007). Nonetheless, the observed genomic instability of 6q might result in structural alterations promoting *CTGF* expression that may have been overlooked in genome wide scans. These structural alterations may exist as discreet CNAs or as complex translocations involving the *CTGF* locus. Analysis of *CTGF* expression in the PMH cohort demonstrated that *CTGF* expression was highest in leukaemic blasts containing a recurrent cytogenetic rearrangement in pre-B ALL, with the exception of those harbouring the *TCF3-PBX1* rearrangement (Boag et al. 2007). This indicates that genome instability is possibly associated with transcriptional activation of *CTGF*. Therefore the structure of the *CTGF* locus at 6q23.1 should be examined to establish whether such alterations affect this region directly.

A comprehensive assessment of the DNA sequence and gross genomic structure of the *CTGF* locus was undertaken to identify genomic lesions associated with *CTGF* overexpression. The specific aims of the experiments presented in this chapter were as follows: 1) To characterise *CTGF* expression in a panel of B-lineage ALL cell line models, 2) to examine these ALL cell lines for promoter and 3'UTR mutations that may contribute to *CTGF* overexpression, and 3) to evaluate the *CTGF* genomic locus for structural rearrangements or copy number alterations.

4.2 Results

4.2.1 Cell line models of paediatric B-lineage ALL

Cell lines represent a valuable tool for the molecular modelling of human diseases, as they provide an inexhaustible supply of proteins and nucleic acids for analysis of tumour biology, as well as facilitating functional studies *in vitro*. Numerous B-lineage ALL cell lines have been generated from primary samples in the laboratory of Prof. Kees. These

cell lines were derived by serial passage of paediatric patient specimens *in vitro* and experiments have confirmed they mirror critical features of the primary patient specimens from which they were derived (Kees et al. 2003; Beesley et al. 2006). The cell lines that were chosen as a model for B-lineage ALL in this study are shown in Table 4.1. They include three pre-B ALL, two B-ALL and two infant-ALL cell lines with B-lineage features. All of the cell lines have complex karyotypes and all harbour genetic features known to contribute to leukaemogenesis, namely; *TEL-AML1*, *TCF3-PBX1* and *MYC-IgH* translocations as well as rearrangements involving the *MLL* locus at 11q23. The *CTGF* locus is located on the long arm of chromosome 6 at cytoband 6q23.1 and there have been no alterations of this locus observed in any of these cell lines to date.

Table 4.1 Features of B-lineage ALL cell lines

Cell Line	Subtype [§]	D/R [‡]	Karyotype	Genetic Feature	Refs
PER-145	pre-B	R	45,XY, der(3)t(3;?)(q12;?), der(4)t(4;?)(p15.2;?), t(5;17)(q15;p13), der(7)t(7;8)(p13-14;q13-21), -8,?t(9;20;9)(p13;p12;q34), der(12)t(12;?)(p13;?)	<i>TEL-AML1</i>	(Kees 1987)
PER-278	pre-B	D	46,XY,der(9)t(1;9)(q23;p13),der(19)t(1;19)(q23;p13)	<i>TCF3-PBX1</i>	(Kees et al. 1990)
PER-371	pre-B	D	46,XY, der(16)t(1;16)(q2?1;p13), der(19)t(1;19)(q?13;p13) 46,X,-Y,+?der(1)t(Y;1)(q12;?q21), add(11)(q21), der(19)t(1;19)	<i>TCF3-PBX1</i>	(Kees et al. 2003)
PER-377	B-ALL	R	46,XY, t(2;13)(p12;q34), del(7)(q11q21), ?inv(14)(q11q24), der(17)t(8;17)(q11;p11)	cryptic <i>MLL-AF9</i>	(Kees et al. 1995; Whitman et al. 2001)
PER-495	B-ALL	R	46,XY,t(8;14)(q24;q32)	<i>MYC-IgH</i>	(Kees et al. 2003)
PER-485	infant-ALL	R	47,XX, der(4)t(4;11)(q21;q23) add(4)(p16),_6, del(7)(p14), add(8)(q24.3), der(9)inv(9)(p11q12) del(9)(p24), der(11)t(4;11)(q21;q23)	<i>MLL</i> rearranged	(Kees et al. 2003)
PER-490	infant-ALL	D	46,XX,t(4;11)(q21;q23) 46,XX,t(4;11),dup(1)(q12q44) 46,XX,t(4;11),der(2)t(1;2)(q12;q37)	<i>MLL</i> rearranged	(Kees et al. 2003)

§ Subtype: Classification of ALL specimens by immunophenotype.

‡ Specimens obtained at diagnosis (D) or relapse (R)

4.2.2 *CTGF* expression in B-lineage ALL cell lines

To gain an insight into the range of *CTGF* mRNA levels in these cell lines, expression was measured in seven B-lineage ALL cell lines and one commercial T-ALL cell line (JURKAT) by qRT-PCR. Those cell lines that expressed detectable levels of *CTGF* mRNA were classed as *CTGF*^{high} or *CTGF*^{low}. As shown in Figure 4.1, PER-377 expressed *CTGF* mRNA at levels approximately five and a half times that of the next highest cell line, PER-145 and was classed as *CTGF*^{high}. PER-145, PER-278 and PER-371 expressed *CTGF* at more modest levels and were classed as *CTGF*^{low}. The remaining cell lines; PER-485, PER-490 and PER-495, as well as the T-ALL cell line JURKAT, were negative for *CTGF* mRNA. All of these cell lines have been analysed previously by microarray and the *CTGF* expression data generated by qRT-PCR were in good agreement with this data.

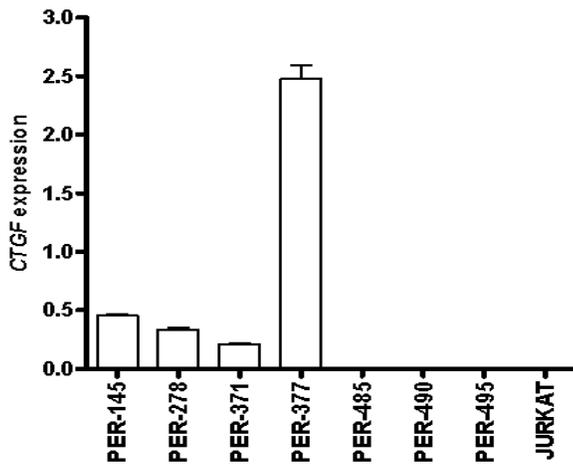


Figure 4.1 *CTGF* gene expression in ALL cell lines

CTGF expression normalized to *ACTB* was measured by qRT-PCR in seven B-lineage ALL and one T-ALL cell line (JURKAT). Error bars represent the standard error of the mean of three technical replicates.

4.2.2.1 CTGF is secreted by PER-377

CTGF is a secreted protein and may have important extracellular roles in signalling between pre-B ALL cells and the bone marrow microenvironment. B-lineage ALL cell lines were examined for CTGF protein synthesis by immunoblotting using two antibodies outlined in Figure 4.2A. These were the L20 antibody which recognised an epitope encompassing the C-terminal region of module 1 and the N-terminal region of module 2 (amino acids 150 to 200) of the CTGF protein, and the polyclonal Rabbit-A antibody which recognised the C-terminus of the CTGF protein between amino acids 247 and 260 (Brigstock et al. 1997). Whole protein extracts (75µg) from the three cell lines that exhibited the highest *CTGF* mRNA levels; PER-377, PER-145 and PER-278 were examined by immunoblotting using the L20 antibody with the T-ALL cell line JURKAT serving as a negative control. Recombinant human CTGF (rhCTGF), which was generously provided by Prof. David Brigstock, was used as a positive control.

Full-length CTGF (38kDa) was detected in extracts from PER-377 as shown in Figure 4.2B, which was classed as *CTGF*^{high} by qRT-PCR. No CTGF protein was detected in the *CTGF*^{low} cell lines PER-145 or PER-278. CTGF protein may be produced in these cells below the level of detection by immunoblotting, or alternatively the protein may be rapidly secreted making it difficult to detect in whole protein extracts. Efforts were therefore directed towards detecting secreted CTGF in cell culture supernatants. Conditioned cell culture medium (CM) from the *CTGF*^{high} cell line PER-377 was examined using the Rabbit-A antibody which was a gift from Prof. David Brigstock and has been used previously to detect CTGF proteins in biological fluids (Brigstock et al. 1997). Cells were seeded in 1ml of culture medium, containing 20% fetal calf serum (FCS) denoted as A+20, at a density of 10^6 cells/ml and grown for 24 hrs, after which time CM was collected and pooled for analysis.

Immunoblotting of 50µl of unprocessed (Raw) CM or untreated culture medium (A+20), is shown in lanes 1 and 2 of Figure 4.2C. These lanes appeared overloaded and there was excessive cross-reactivity with the Rabbit-A antibody resulting in a failure to adequately detect secreted CTGF. This cross-reactivity was attributed to the high levels

of FCS used to culture the cell lines, as FCS contains a high concentration of bovine immunoglobulins. To overcome this, a heparin-affinity purification method was developed using heparin-conjugated agarose beads. Briefly, CM harvested at 24 hrs was pooled to a volume of 5ml, then incubated with heparin-agarose beads for 24 hrs. After washing beads, the bound protein was eluted directly into 200µl of hot SDS-PAGE sample buffer. The CTGF protein contains a heparin binding motif in the CT domain and enriching supernatants by this method improved the detection of CTGF from conditioned medium greatly as shown in lanes 3 and 4 of Figure 4.2C.

This approach was used to screen heparin-enriched CM from eight ALL cell lines using the L20 antibody, as immunoblots performed with this antibody were of superior quality compared to those using the Rabbit-A antibody. Nevertheless, when the entire panel of cell lines was screened for secreted CTGF (shown in Figure 4.2D), only the *CTGF*^{high} cell line PER-377 secreted detectable CTGF, evident as a 36-38kDa doublet which represented the differentially glycosylated forms of the full length protein (Ball et al. 2003a). From these results it was clear that *CTGF* mRNA levels did not reflect protein levels *in vitro*. This disparity may simply be due to the high level of CTGF mRNA expression in PER-377 compared to the *CTGF*^{low} cell lines, however post-transcriptional mechanisms may also be involved in regulating CTGF protein synthesis in these cell lines.

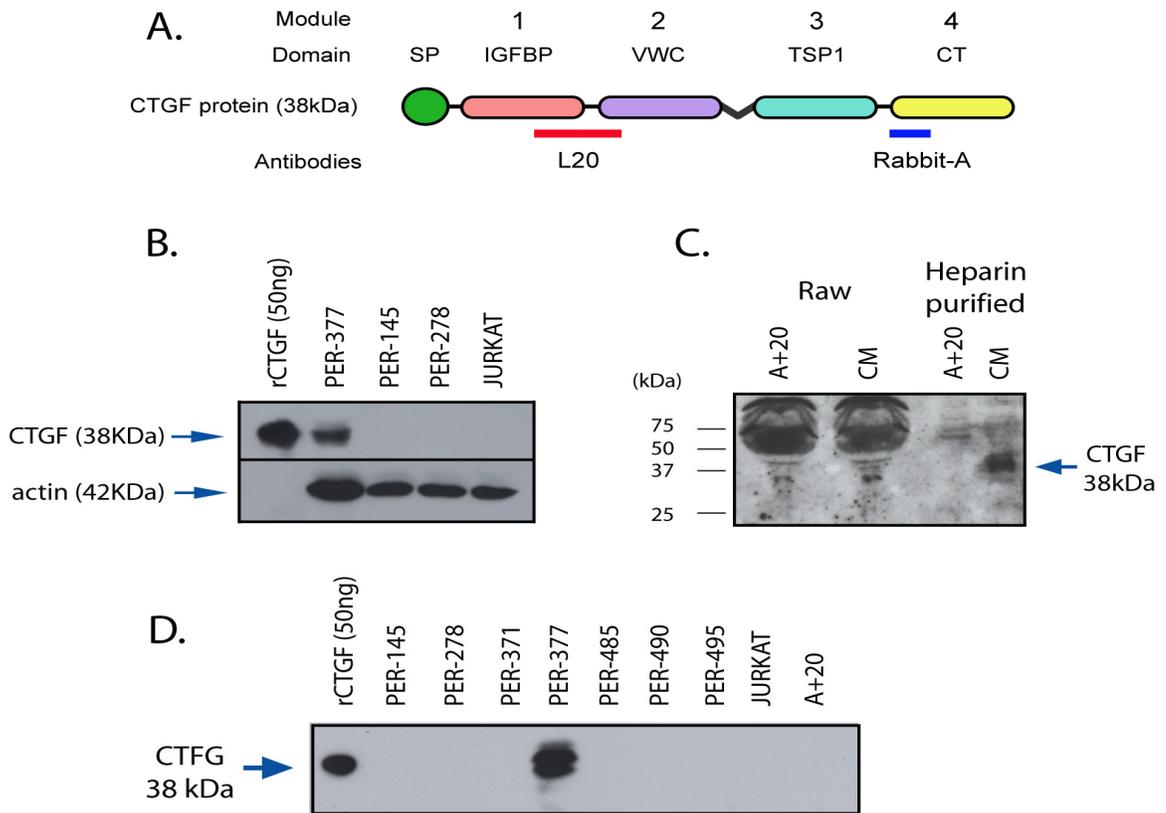


Figure 4.2 Detection of CTGF protein by immunoblotting

A. Structure of the CTGF protein and epitopes targeted by the L20 and Rabbit-A antibodies. **B.** Immunoblotting of 75 μ g of total protein performed with the L20 antibody. 50ng of Recombinant CTGF (rCTGF) was used as a positive control. **C.** Immunoblotting of 50 μ l of A+20 or CM from PER-377 without heparin purification (Raw) or after heparin purification using the Rabbit-A antibody. **D.** Detection of CTGF protein in 50 μ l of heparin-purified CM from seven B-lineage ALL and one T-ALL cell line or control media (A+20) using the L20 antibody.

4.2.3 Analysis of *CTGF* 5' flanking sequence

Sequencing of the *CTGF* promoter and 5' flanking sequence was undertaken to determine if promoter mutations may contribute to dysregulation of *CTGF* expression in ALL. Three cell lines were chosen as appropriate models with which to screen the *CTGF* promoter for mutations. These were the *CTGF*^{high} cell line PER-377 and the

CTGF^{low} cell lines PER-145 and PER-278. Sequencing of the *CTGF* promoter was undertaken using three separate overlapping PCR amplicons. Figure 4.3 outlines the *CTGF* promoter and 5' flanking sequence. Validated transcription factor-binding sites are highlighted, as well as the predicted binding site for Ikaros family proteins identified in Chapter 3. Placement of primers used to amplify regions of interest and sequence PCR products are also shown. PCR products were gel extracted, purified and sequenced directly in both forward and reverse orientation and aligned to the human chromosome 6 reference sequence.

Sequencing provided complete coverage of the *CTGF* promoter in all three cell lines and high quality sequencing reads were generated in both forward and reverse orientations. The only observed deviation from the reference sequence was found in the PER-278 cell line. This cell line was heterozygous for the rs6918698 C>G SNP at position -739 as shown in Figure 4.4A. The C>G transversion is thought to promote *CTGF* expression by disrupting repression by the SP3 nuclear factor (Fonseca et al. 2007). To determine if this polymorphism may have other functional impacts, analysis of the flanking sequence containing either the C or G allele was performed using the PROMO transcription factor binding prediction program (Farre et al. 2003; Messeguer et al. 2002) and the TFSEARCH program applied in Chapter 3. Putative transcription factor-binding sites are shown in Figure 4.4. Despite the findings by Fonseca and colleagues in relation to Sp3-mediated repression of *CTGF* expression in airway fibroblasts, neither the C nor G allele was predicted to facilitate transcription factor binding by PROMO and TRANSFAC. Nevertheless, this site has been functionally validated as a binding site for SP3 in airway fibroblasts (Fonseca et al. 2007). Further investigation into the incidence of the rs6918698 -739G genotype in *CTGF* positive specimens may be warranted.

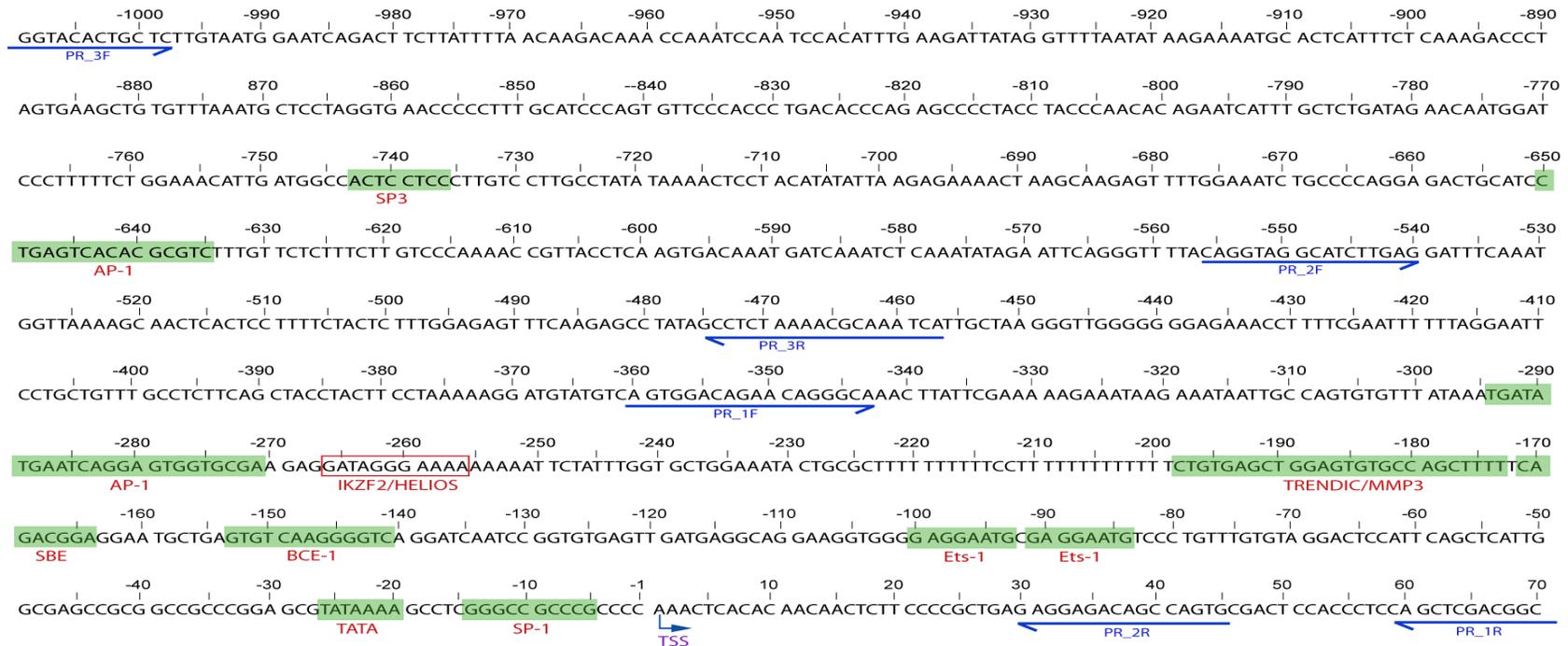


Figure 1.3 Transcription factor binding sites at the *CTGF* promoter

Primers used to generate the three sequencing amplicons (PR_1-3) from the *CTGF* promoter (-1009 to +70) are indicated by blue arrows. Validated transcription factor binding sites (green boxes) include two AP-1 sites (Yu et al. 2009), transcriptional enhancer dominant in chondrocytes (TRENDIC) which is bound by MMP3 (Eguchi et al. 2007; Eguchi et al. 2008), an SP3 binding site (Fonseca et al. 2007), a Smad binding element (SBE) (Arnott et al. 2008), basal control element (BCE-1) (Grotendorst 1997), tandem Ets-1 binding sites (Van Beek et al. 2006), a TATA box and an SP-1 binding site (Holmes et al. 2003). In addition, the location of the predicted HELIOS binding site is indicated by the red open box. TSS: transcription start site (+1).

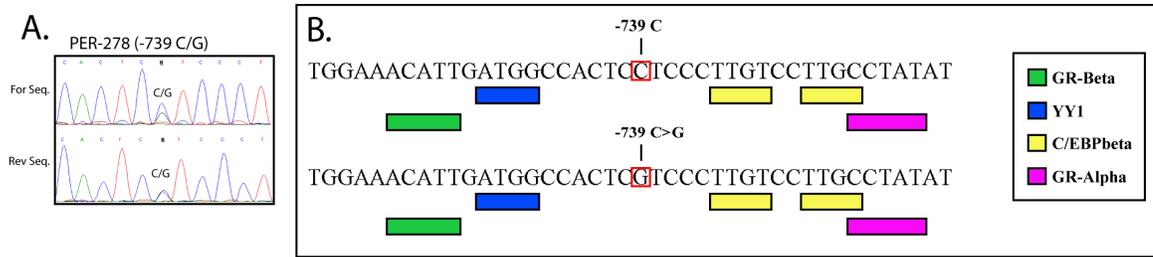


Figure 4.4 Analysis of the rs6918698 C>G SNP in PER-278

Forward and reverse sequencing of the *CTGF* promoter in PER-278 found this cell line to be heterozygous for the 6918698 C>G SNP at position -739. **B.** This SNP, highlighted by an open red box, together with flanking sequence, was examined for transcription factor binding motifs using the PROMO and TFSEARCH transcription factor binding prediction programs. Predicted TF binding motifs are indicated by coloured boxes. None of the identified motifs were associated with the rs6918698 SNP.

4.2.4 Sequencing of *CTGF* 3' UTR

The discrepancy between *CTGF* mRNA and protein levels in *CTGF*^{low} cell lines suggests some level of post-transcriptional regulation affecting the translational potential or longevity of *CTGF* mRNA. The *CTGF* 3'UTR contains several experimentally validated miRNA binding sites (Ohgawara et al. 2009; Duisters et al. 2009; Ernst et al. 2010) and mutation of miRNA binding sites may result in increased expression *via* inhibition of miRNA binding or alternatively, the introduction of a new miRNA binding site by mutagenesis may result in translational repression or increased turnover of mRNA. The miRNA target prediction software TargetScan available online at <http://www.targetscan.org/> (Friedman et al. 2009) was used to screen the *CTGF* 3'UTR for predicted miRNA binding motifs that may be of interest in addition to known binding sites. Figure 4.5 illustrates the known miRNA targets in the *CTGF* 3' UTR that have been functionally validated, including those for miR30, miR133, miR18a, as well as sites predicted by TargetScan, including binding sites for miR26a/1297, miR-132/212 and miR-19. Three PCR amplicons were generated for the 3'UTR as outlined in Figure 4.5. Primers pairs used to generate these amplicons were as follows, amplicon 1; CD_7F

& CD_7R, amplicon 2; CD_8F & CD_8R, and amplicon 3; CD_9F & CD_9R. Sequencing of these amplicons produced high quality reads in forward and reverse orientation for the PER-145, PER-278 and PER-377 cell lines. Comparing these sequences to the chromosome 6 reference sequence did not reveal any 3' UTR mutations in these cell lines. This confirmed that all known miRNA binding sites were intact in the *CTGF* 3'UTR.

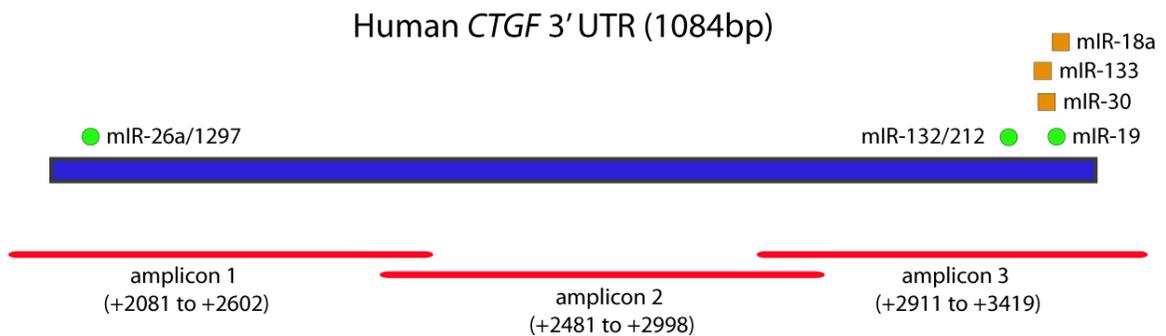


Figure 4.5 miRNA binding sites in the *CTGF* 3' UTR

The *CTGF* 3' UTR (blue) is 1084 bases in length. Predicted miRNA binding sites are indicated by green circles. Functionally validated miRNAs are shown by orange squares. Red lines (bottom) indicate overlapping PCR amplicons used to sequence the *CTGF* 3' UTR. Coordinates are relative to the *CTGF* transcription start site.

4.2.5 Structure of the *CTGF* gene locus

Besides single-base mutations, larger scale genomic rearrangements can result in activation of proto-oncogenes by juxtaposing coding regions adjacent to active promoters or similarly by moving enhancer or promoter elements to a normally silent coding region (Pui 2006). Although no 6q23 rearrangements were evident by karyotype analysis of the B-lineage ALL cell lines (shown in Table 4.1), rearrangements such as gene copy number alterations or cryptic translocations affecting the *CTGF* gene locus may influence *CTGF* gene transcription in pre-B ALL.

To examine whether the *CTGF* locus is frequently affected by genomic rearrangements in ALL, the Mitelman database of chromosomal aberrations and translocations in cancer (Mitelman et al. 2010) was screened for alterations involving 6q23.1. The results of this analysis are summarised in Table 4.2. There have not been any recorded focal amplifications of 6q23.1 in ALL, however there has been numerous cases involving deletion of regions both upstream and downstream from 6q23, indicating that this region has a relatively high frequency of rearrangement (Mitelman et al. 2010). Furthermore, data from the Mitelman database suggested that trisomy of chromosome 6 is nine-fold higher than deletion of chromosome 6 in ALL.

Table 4.2 Documented aberrations involving 6q23 in ALL

Structural alterations involving the 6q23 locus (left panel) and gross numerical alterations involving chromosome 6 (right panel) in ALL. Source: Mitelman database of chromosomal aberrations and translocations in cancer (Mitelman et al. 2010)

6q23 Alterations			Numerical Aberrations	
Band	Abnormality	Cases		
6q23	del(6)(q13q23)	18	Trisomies	
6q23	del(6)(q14q23)	4	Abnormality	Cases
6q23	del(6)(q15q23)	29	+6	892
6q23	del(6)(q16q23)	4		
6q23	del(6)(q21q23)	25	Monosomies	
6q23	del(6)(q23)	31	Abnormality	Cases
			-6	102
6q23	del(6)(q23q25)	15		
6q23	del(6)(q23q26)	2		
6q23	del(6)(q23q27)	4		

Analysis of the expression of genes proximal to the *CTGF* coding region was performed using global gene expression data from the PMH cohort described in Chapter 3, to determine if gene expression in this region was broadly deregulated or followed a similar pattern to *CTGF*. Six genes in total were analysed. These included genes on the centromeric side of the *CTGF* coding region at 6q23.1, namely A-Kinase anchor protein 7 (*AKAP7*), mediator complex subunit 23 (*MED23*) and ectonucleotide pyrophosphatase/phosphodiesterase 1 (*ENPP1*), and three genes on the telomeric side of the *CTGF* locus, namely syntaxin 7 (*STX7*), trace amine associated receptor 5 (*TAAR5*)

and *vanin3* (*VNN3*). When expression of these genes was compared between *CTGF* positive and low/negative samples, there was no significant difference in expression between the two groups, as shown in Figure 4.6. Based on these data it is unlikely that deregulation of the 6q23 region as a whole contributes to *CTGF* overexpression in pre-B ALL.

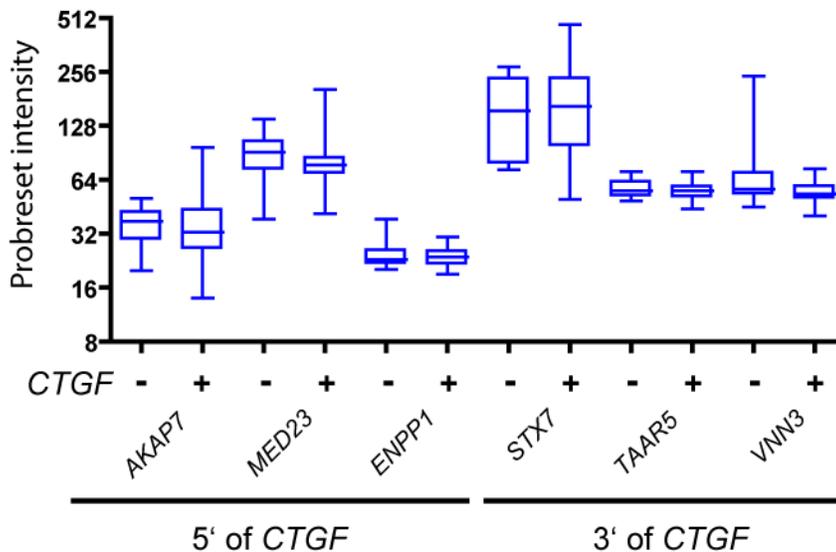


Figure 4.6 Expression of genes neighbouring the *CTGF* coding region

Expression of six genes in close proximity to the *CTGF* coding region at 6q23.1 measured in the PMH cohort by microarray. Genes were selected based on their position relative to the *CTGF* coding region. *AKAP7*; 205771_s_at, *MED23*; 218846_at and *ENPP1*; 205066_s_at are located on the centromeric side of the *CTGF* coding region, while *STX7*; 212631_at, *TAAR5*; 221459_at and *VNN3*; 220528_at are located on the telomeric side of the *CTGF* gene. For each gene, expression levels are shown in *CTGF*^{low/neg} (minus symbol, n=13) or *CTGF*^{med/high} (plus symbol, n=60) samples.

4.2.5.1 Southern blotting of the *CTGF* locus

There is evidence that internal deletions affecting the long arm of chromosome 6 occur at a relatively high frequency in pre-B ALL (Thelander et al. 2008). Genome instability may result in translocations that have profound effects on gene expression. To

investigate whether the *CTGF* locus is affected by genomic rearrangements, Southern blotting was performed on seven B-lineage ALL cell lines. The hybridisation probe and restriction sites used are outlined in Figure 4.7. Genomic DNA was digested with either Hind III or Psi I, gel fractionated and hybridised with a 4.4kb probe spanning the *CTGF* locus from -1024 to +3418 relative to the transcription start site (TSS). Digestion with Hind III was predicted to yield two fragments; 12.5kb and 2kb, capable of being hybridised by the *CTGF* probe, while digestion with Psi I produced two fragments; 2.6kb and 6.5kb. These two digests provided coverage of the *CTGF* locus from nucleotide position -9660 to +6224. Figure 4.8 shows the result of Southern blotting performed on digested genomic DNA from seven B-lineage ALL cell lines and one PBMC control sample from a healthy donor (C5). Fragments of the expected size were observed in all cases with no extraneous bands identified. Thus, there were considered to be no detectable rearrangements of the *CTGF* locus present in these samples.

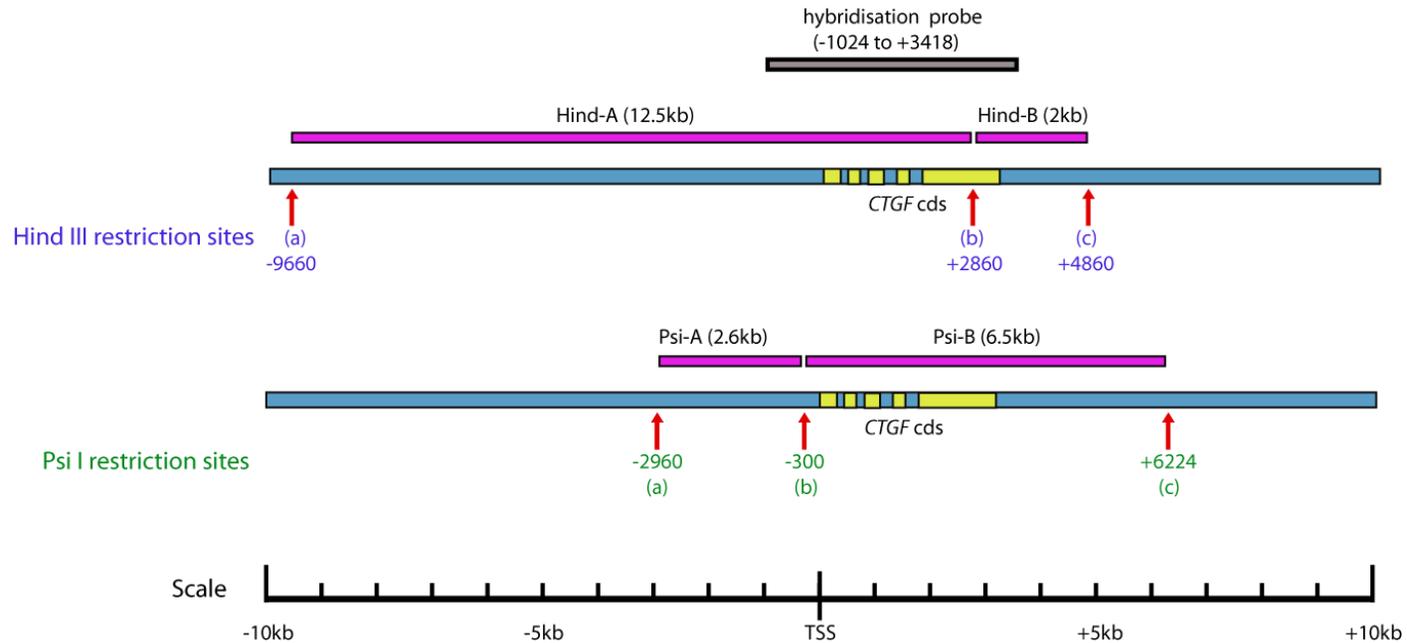


Figure 4.7 Localisation of hybridisation probe and restriction sites used for Southern blotting analysis of the *CTGF* locus
 10kb either side of the *CTGF* transcription start site (TSS) is shown above. Sequence flanking the *CTGF* coding sequence (cds) is shown in blue, while exons are indicated by yellow boxes. Scale bar at the bottom indicates distance in kilobases. Restriction sites for Hind III (top) are denoted by red arrows at -9660 (a), +2860 (b) and +4860 (c), while restriction sites for Psi I (bottom) occur at -290 (a), -300 (b) and +6224 (c). A 4.4kb *CTGF* probe (grey bar at top) spanning from -1024 to +3418 hybridises digest fragments shown as purple bars; Hind-A (12.5kb) and Hind-B (2kb) as well as two Psi I fragments; Psi-A (2.6kb) and Psi-B (6.5kb).

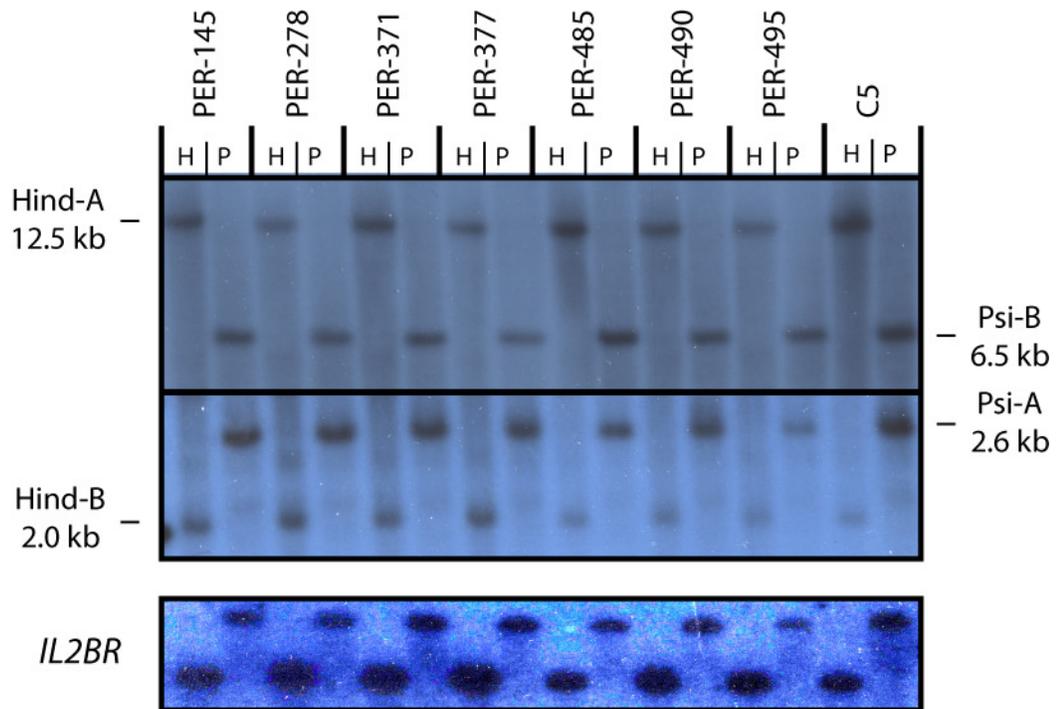


Figure 4.8 Southern blot analysis of the *CTGF* locus

Seven B-lineage ALL cell lines and PBMC gDNA from a healthy donor (C5) analysed by Southern blotting. Samples were digested with either Hind III (H) or Psi I (P) and hybridised to a cDNA probe specific for the *CTGF* locus (see Figure 4.7). Different exposures are shown for optimal resolution of large (top panel) and small fragments (middle panel). The top panel contains large fragments (Hind-A and Psi-B), while middle panel contains smaller MW fragments (Hind-B and Psi-A). Bottom panel contains samples re-probed for the *IL2BR* gene as a loading control.

4.2.5.2 Gene copy number analysis

Southern blotting did not indicate any overt rearrangements of the *CTGF* locus, however there may be CNAs outside of the region detected by Southern blotting. To further examine CNAs, a commercial qPCR gene copy number assay (Applied Biosystems) (*CTGF*:hs00517234_cn) was used to examine 7 B-lineage ALL cell lines and 17 primary pre-B ALL specimens (outlined in Chapter 3). The *CTGF* copy number assay was performed in a duplex reaction with the reference copy number assay targeting the RNase P (*RPPHI*) coding region at 14q11. Assays were run in duplex reactions in two independent experiments and data was imported into the CopyCaller software package (Applied Biosystems) for analysis. Samples were considered to contain a copy number alteration if they crossed a threshold of 3.0 indicated on the Y axis in Figure 4.9. The reference PBMC DNA samples A4 and A5 were isolated from healthy donors and used as diploid calibrator samples for these experiments. Analysis of the data using the CopyCaller software predicted alterations in *CTGF* copy number in two cell lines; PER-377 (amplification) and PER-495 (deletion).

Upon manual inspection of the raw copy number assay data, it was noted that there was significant variation in the cycle threshold (C_t) values obtained for the RNaseP control assay in cell lines, which may confound interpretation of *CTGF* copy number. Figure 4.10A shows the raw C_t values generated by both assays from the B-lineage ALL cell lines. The *CTGF* assay C_t values suggested that only PER-495 contained a genuine alteration in *CTGF* copy number, namely a deletion of one *CTGF* allele. Data obtained from the RNaseP assay suggested that PER-377 and potentially PER-490 may contain less than two copies of the RNaseP locus. Furthermore, RNaseP is located at 14q11.2 and the cytogenetics of PER-377 indicates that this cell line harbours a cryptic rearrangement on chromosome 14 between q11 and q24.

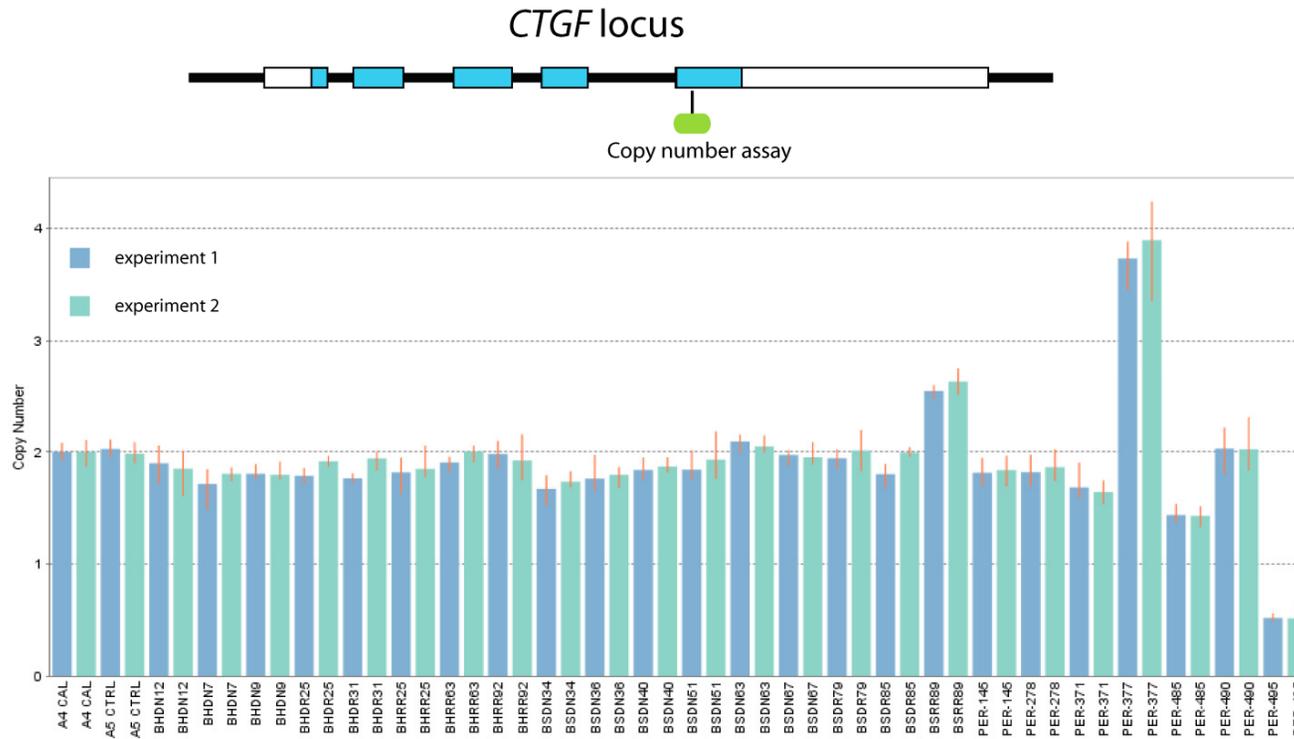


Figure 4.9 *CTGF* copy number analysis

Top; the *CTGF* copy number assay (hs00517234_cn) targets exon 5 within the *CTGF* coding region. Bottom; results of the copy number assay from 2 independent experiments normalised against the reference assay for ribonuclease P RNA component H1 (*RPPH1*). For each experiment, error bars represent the standard error of the mean of technical replicates (n=4). A4 and A5 gDNA was obtained from PBMC isolated from healthy individuals.

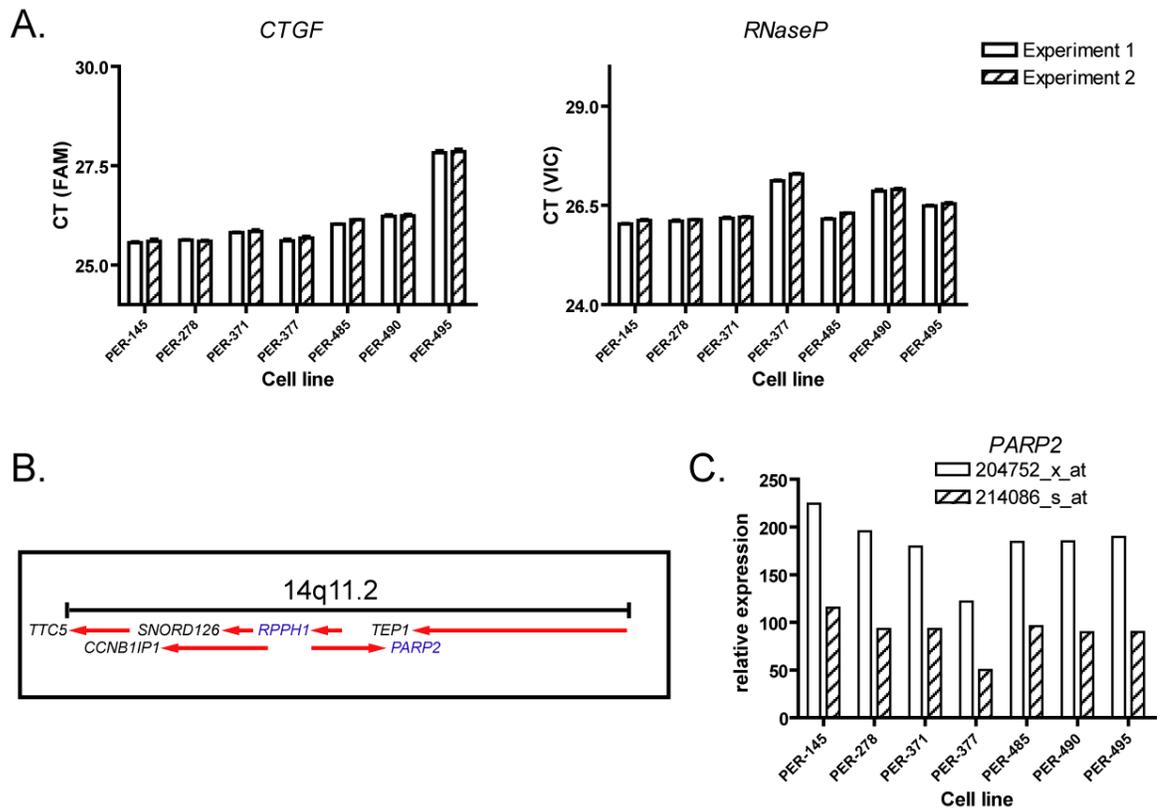


Figure 4.10 Raw gene copy number data obtained from ALL cell lines

A. Taqman copy number assay data generated from the *CTGF* assay (left) and the reference *RNaseP* assay (right). Raw (non normalized) C_t values are plotted for each of the seven B-lineage ALL cell lines. Two biological replicates (experiments 1 and 2) were conducted. Error bars represent the standard error of the mean of technical replicates ($n=4$). **B.** The *RNaseP* copy number reference assay targets the coding region of *RNaseP* (*RPPH1*) at 14q11.2. The *PARP2* gene is located on the opposite strand at this locus. **C.** *PARP2* expression measured by microarray (204742_x_at, 214086_s_at) in seven B-lineage ALL cell lines.

Gene expression at the 14q11.2 locus was examined, as loss of one allele should be reflected in changes in expression of affected genes. There was no Affymetrix probeset targeting *RNaseP* (*RPPH1*) on the U133A platform, however there is a probeset for *PARP2*, which is located in close proximity to *RPPH1* on the opposite genomic strand as shown in Figure 4.10B. The expression level of *PARP2* was determined using microarray profiles from our B-lineage ALL cell lines and it was found that there was

indeed a reduced expression of the two *PARP2* probesets (204752_x_at, 214086_s_at) in PER-377 (Figure 4.10-C), supporting the hypothesis that PER-377 harbours a deletion at the *RPPH1/PARP2* locus. Based on this finding as well as the observed differences in C_t values for the *RNaseP/RPPH1* copy number assay, the most likely explanation is that PER-377 contains two copies of *CTGF* and harbours a deletion of *RPPH1*, while PER-495 contains a focal deletion of one copy of *CTGF*. All primary specimens tested were diploid for *CTGF*.

4.3 Discussion

Alterations in DNA sequences are considered to be the earliest events underpinning neoplastic transformation of pre-cancerous cells (Greaves 2009). Recurrent chromosomal translocations and genome instability are a feature of leukaemia cells, thus determining if overexpression of *CTGF* mRNA occurs in the absence of DNA rearrangements or mutations was considered an important step in elucidating the molecular mechanisms contributing to *CTGF* deregulation. Analysis of the *CTGF* locus was carried out in a panel of seven B-lineage ALL cell lines, which had been previously evaluated for *CTGF* expression (Kees et al. 2003; Beesley et al. 2006). Further analysis of *CTGF* expression levels in these cell lines revealed that only the *CTGF*^{high} cell line PER-377 synthesised CTGF protein. This suggested that post-transcriptional regulation may repress CTGF protein synthesis in the *CTGF*^{low} cell lines. A study into the prognostic significance of *CTGF* mRNA expression in adult ALL observed a poorer outcome in patients that expressed *CTGF* mRNA at high levels in their leukaemic blasts compared to those with low or intermediate level expression (Sala-Torra et al. 2007), thus there may be a threshold below which CTGF protein synthesis is repressed post-transcriptionally. Stratifying patients in the PMH cohort by *CTGF* mRNA expression (high, medium or low) was not predictive of RFS or outcome (data not shown), thus careful immunological examination of CTGF protein expression in trephine bone marrow biopsies should be performed and coupled with qPCR analysis from matched leukaemic blasts, to determine if such a threshold exists as this may have significant prognostic relevance for pre-B ALL patients.

The 3' UTR in exon 5 of *CTGF* contains several miRNA binding sites that have been experimentally validated, as well as a number of predicted binding sites identified in the present study (Duisters et al. 2009; Ohgawara et al. 2009). Sequencing this region in three *CTGF* positive cell lines confirmed that no mutations affected the 3' UTR, thus full-length, canonical *CTGF* transcripts should be amenable to post-transcriptional regulation by miRNAs. Profiling of miRNA expression levels in pre-B ALL was considered beyond the scope of the present study, however post-transcriptional regulation by miRNAs may be responsible for the apparent lack of CTGF protein in *CTGF*^{low} cell lines. Global miRNA profiling integrated with gene expression data would provide some clues as to the role of miRNAs in regulating *CTGF* and other genes in pre-B ALL (Lionetti et al. 2009; Volinia et al. 2010). Furthermore, inhibiting specific miRNAs such as those already validated as *CTGF* regulators, in *CTGF*^{low} cell lines may confirm a role for these miRNAs in *CTGF* expression in pre-B ALL.

4.3.1 *CTGF* promoter mutation was not detected in pre-B ALL

Mutation of the *CTGF* promoter or 5' flanking sequence was considered a possible mechanism by which *CTGF* overexpression may be potentiated in pre-B ALL. Analysis of the *CTGF* locus confirmed that the promoter and 5' flanking region from pos -1009 to +70 did not contain any DNA sequence mutations. However, the PER-278 cell line harboured the rs6918698 C>G SNP at -739. This polymorphism is contained within a binding site for the ubiquitously expressed nuclear factor SP3, and transversion from C to G can disrupt SP3-mediated repression of the *CTGF* locus (Fonseca et al. 2007). Functional studies conducted by Fonseca and colleagues in primary airway fibroblasts confirmed that SP3 can bind to this site when the -739 C nucleotide is present and repress basal and thrombin-induced *CTGF* promoter activation (Fonseca et al. 2007). SP3 is a dual function transcription factor that can induce or repress gene expression through interaction with different co-factors. SP3 can undergo post-translational modification by small ubiquitin-like modifier (SUMO) proteins, and these modifications promote repression of SP3-bound loci (Stielow et al. 2008). Furthermore, SUMO proteins have been shown to recruit chromatin-modifying enzymes and DNA

methyltransferases, resulting in compact, transcriptionally silent heterochromatin and increased site-specific DNA methylation, thus SP3 may act to target or maintain epigenetic silencing at the *CTGF* locus (Stielow et al. 2010).

Genotyping studies have investigated the association of this SNP with *CTGF* overexpression in systemic sclerosis (SSc), as serum and dermal *CTGF* levels are elevated in patients with this disease (Sato et al. 2000; Igarashi et al. 1995). Some studies have identified significant associations between the rs6918698 G allele and SSc (Fonseca et al. 2007; Kawaguchi et al. 2009), however, other studies have not been able to replicate these associations in independent cohorts (Rueda et al. 2009; Gourh et al. 2008; Granel et al. 2010). While this SNP was only identified in one of the three cell lines that were sequenced, it may have a higher penetrance in primary pre-B ALL specimens. Based on the findings by Stielow and colleagues, germline transversion of this SNP to the G allele may result in a failure of SP3 to target epigenetic repression of the *CTGF* locus and subsequently the *CTGF* locus would be permissive to improper transcriptional activation (Stielow et al. 2010). Previous studies investigating frequent genetic alterations in pre-B ALL have employed genome-wide SNP-array technology (Kuiper et al. 2007b; Mullighan et al. 2007), however a probe for the the rs6918698 SNP is not contained on Affymetrix array platforms and thus was not interrogated in these studies. Future experiments are justified to undertake genotyping of this SNP in a cohort of primary pre-B ALL specimens, and to establish if there is an association between the rs6918698 SNP and *CTGF* gene expression more broadly in ALL.

4.3.2 The *CTGF* locus in not structurally altered in pre-B ALL

In addition to single base DNA mutations, genomic translocation or gene amplification may be associated with *CTGF* overexpression. Genome wide scans have identified frequent internal deletions in the long arm of chromosome 6, however the *CTGF* locus remains largely unaffected by such alterations (Thelander et al. 2008; Mullighan et al. 2007). This finding was confirmed by analysis of the Mitelman database of Chromosome Aberrations and Gene Fusions in Cancer which contains cytogenetic

information from 58,819 primary cancers (Mitelman et al. 2010). Analysis of 6q alterations in ALL revealed that areas adjacent to *CTGF* at 6q23.1 are frequently affected by deletions. This genetic instability may result in translocation or rearrangement of the *CTGF* locus, or the juxtaposition of enhancer elements from distant loci close to the *CTGF* promoter. Analysis of gene expression in the PMH cohort revealed that expression of genes adjacent to *CTGF* at 6q23 was independent of *CTGF* expression, ruling out amplification or any kind of positional effect upon 6q23 as a whole. Southern blotting confirmed that the *CTGF* locus was not rearranged in the ALL cell lines studied, and direct copy number analysis of the *CTGF* coding region by qPCR ruled out amplification of the *CTGF* coding region in seven ALL cell lines and 17 primary pre-B ALL specimens. These findings confirmed that structural alterations of the *CTGF* locus do not contribute to *CTGF* overexpression in pre-B ALL. Gene copy number analysis indicated that the *CTGF*^{high} cell line PER-377 was diploid for *CTGF*.

4.3.3 Conclusions

In summary, the experiments presented in this chapter identified a disparity between *CTGF* mRNA and protein levels suggesting that post-transcriptional regulation of *CTGF* mRNA may occur in *CTGF*^{low} cell lines. This finding highlights the need to evaluate leukaemic blasts for CTGF protein using immunological techniques, as qRT-PCR alone may not be sufficient to gauge the CTGF protein status of leukaemic cells. Analysis of the *CTGF* locus in B-lineage ALL cell lines and primary patient specimens ruled out structural alterations in the *CTGF* locus as contributing to *CTGF* expression in pre-B ALL. Furthermore, the 5' flanking sequence and 3' UTR were wild-type, confirming that promoter and 3' UTR mutation is not a mechanism by which *CTGF* expression is promoted in pre-B ALL. Finally, this study highlights the potential importance of the rs6918698 C>G SNP in the repression of the *CTGF* locus and suggests that further analysis of this SNP by genotyping in patient cohorts is warranted.

Chapter 5

Characterisation of *CTGF*

Transcription in B-lineage ALL

Chapter 5

Characterisation of *CTGF* Transcription in B-lineage ALL

5.1 Introduction

Aberrant *CTGF* gene expression in pre-B cell ALL has been identified in a number of studies as previously discussed. These findings have all been made based upon the level of *CTGF* mRNA detected in leukaemic blasts by either qRT-PCR or array based methods (Vorwerk et al. 2000; Sala-Torra et al. 2007; Boag et al. 2007). It is unclear whether *CTGF* mRNA transcribed in leukaemic cells is only represented by the canonical 2.4 kb transcript or whether there may be other *CTGF* transcripts synthesised in these cells that exhibit altered coding potential. This possibility was highlighted by findings presented in Chapter 4 showing that CTGF protein was detectable in only one of four cell lines that were positive for *CTGF* mRNA. While there were large quantitative differences in *CTGF* mRNA levels between the *CTGF*^{high} cell line PER-377 and the three *CTGF*^{low} cell lines PER-145, PER-278 and PER-371, the disparity in CTGF protein expression raises the possibility that there may be non-canonical *CTGF* transcripts synthesised in these cell lines that are either non-coding or encode CTGF protein isoforms that were not able to be detected by the antibodies used in this study.

Alternative splicing of pre-mRNA to generate transcript variants can increase protein coding diversity without increasing genome size (Hallegger et al. 2010). On the other hand, alternative splicing can also result in the synthesis of truncated mRNA transcripts that may be destined for degradation by nonsense-mediated decay (Peltz et al. 1994), or transcripts missing important regulatory elements such as the 3' UTR, thus modulating protein expression. Since original observations of alternative splicing almost 35 years ago, a great number of mRNA transcript variants have been recorded and it is now estimated that around 90% of human multi-exon genes are alternatively spliced (Mortazavi et al. 2008; Hallegger et al. 2010). Alternative splicing of *CCN* mRNAs including *CCN1*, *CCN3* and *CCN4* has been reported by several groups in the last decade (Perbal 2009). Furthermore *WNT1* (*CCN4*) transcript variants have been

observed in numerous tumour types, including gastric carcinoma, chondrosarcoma and hepatoma (Tanaka et al. 2001; Yanagita et al. 2007; Cervello et al. 2004), while alternative *CYR61* (*CCNI*) transcripts have been identified in breast cancer cell lines (Hirschfeld et al. 2009).

Despite evidence supporting alternative splicing in other CCN family members, alternative *CTGF* transcripts documented by sequencing has not been reported. However, there is some evidence for their existence in the form of truncated cDNAs in expressed sequence tag (EST) databases and a recent study examining the exome of thymic tumours (Soreq et al. 2008). In this study, the authors used Affymetrix exon arrays to identify overexpression of *CTGF* in thymoma specimens. Analysis of the signals generated by probes targeting individual *CTGF* exons indicated a proportional increase in exon 5 and a concurrent decrease in exon 2 compared to other *CTGF* exons (Soreq et al. 2008). This is compelling evidence for the existence of alternative *CTGF* transcripts in these thymic tumours. However, this study was conducted using exon arrays, and no further characterisation of *CTGF* transcripts by sequencing was obtained.

Previous characterisation of *CTGF* transcription was performed using either the Affymetrix microarray platform or qRT-PCR (Vorwerk et al. 2000; Sala-Torra et al. 2007; Boag et al. 2007; Kang et al. 2010). These technologies target discreet regions within *CTGF* mRNA and cannot indicate the size or structure of *CTGF* mRNA transcribed in cells under test. The experiments detailed in this chapter sought to examine *CTGF* mRNA transcripts expressed in B-lineage ALL cell lines, and to characterise these by sequencing any novel *CTGF* transcripts that were identified. The specific aims of the experiments outlined in this chapter were as follows: 1) to evaluate the size and range of *CTGF* mRNA transcripts expressed in ALL cell lines, 2) to characterise any suspected *CTGF* mRNA variants by sequencing, 3) to characterise the 5' and 3' ends of *CTGF* mRNA transcripts.

5.2 Results

Analysis of *CTGF* expression in a panel of B-lineage ALL cell lines in Chapter 4 (summarised in Table 5.1 below) demonstrated that the PER-377 cell line displayed the highest level of *CTGF* mRNA expression. This cell line was therefore chosen to examine *CTGF* mRNA transcripts by northern blotting. Three timepoints were examined to determine if *CTGF* expression was growth dependant. The PER-377 cell line was grown to confluency (2×10^6 cells/ml) and then split 1:1 with fresh culture medium. RNA was extracted at 24, 48 and 72 hrs and two independent northern blots were performed using two different probes designed to target distinct regions of the *CTGF* coding region (Figure 5.1A). The central probe hybridised to the 3' half of the coding region, corresponding to exons 3, 4 and 5, while the 3' probe hybridised to the exon 5 encoded *CTGF* 3' UTR in the same region as the *CTGF* probeset sequence (209101_at) on the Affymetrix human genome array U133A.

Both probes hybridised to the canonical 2.4 kb *CTGF* transcript, as well as to three novel shorter transcripts of approximately 1.8, 1.6 and 1.3 kb in size (Figure 5.1B). While these bands were of a lower intensity, they provided evidence that there were minor *CTGF* transcripts produced in PER-377 and these warranted further investigation. A densitometric analysis (Figure 5.1C) indicated that expression of canonical *CTGF* was highest at 24hrs compared to the later timepoints when normalised to the *ATP5G3* control. Consequently, this timepoint was chosen to analyse the remaining pre-B ALL cell lines by northern blotting in order to determine whether these non-canonical transcripts were also produced in other ALL cell lines.

Table 5.1 *CTGF* mRNA expression in B-lineage ALL cell lines

Expression levels are based upon qRT-PCR values reported in Chapter 4.

B-lineage ALL cell line	<i>CTGF</i> mRNA expression
PER-377	<i>CTGF</i> ^{high}
PER-145 PER-278 PER-371	<i>CTGF</i> ^{low}
PER-485 PER-490 PER-495	<i>CTGF</i> ^{neg}

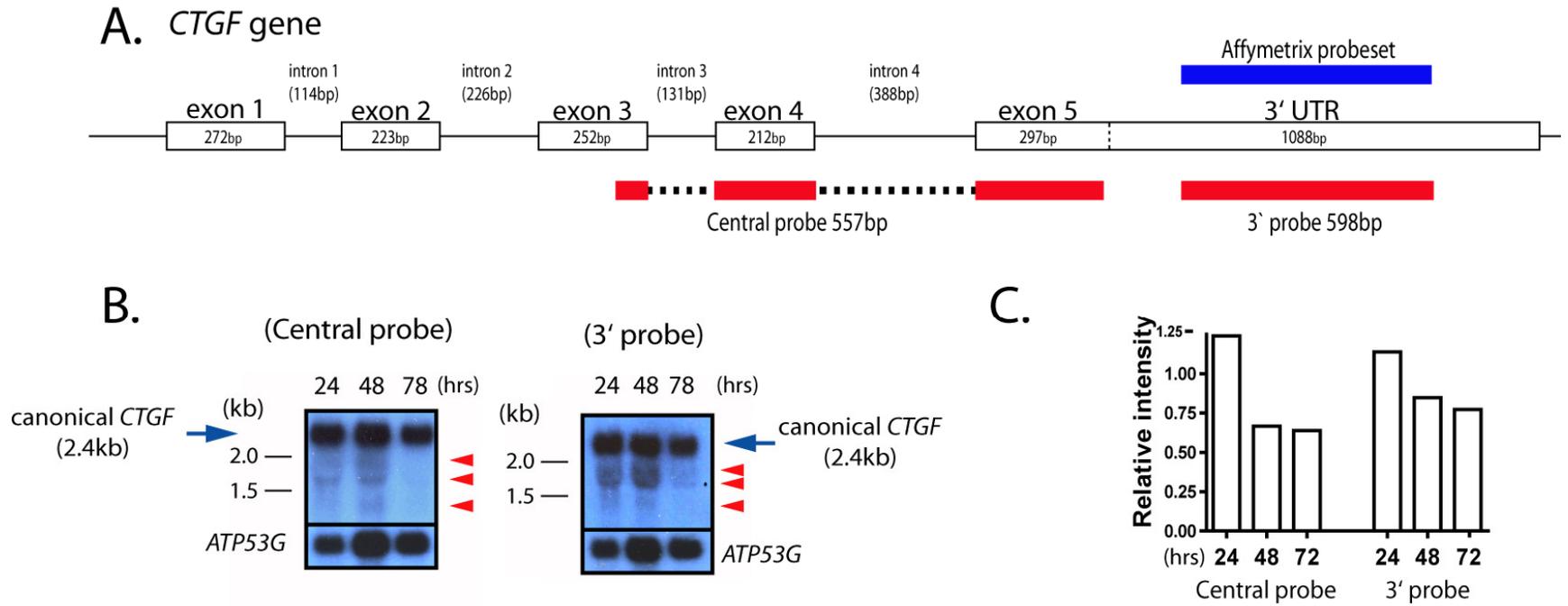


Figure 5.1 Expression of *CTGF* mRNA transcripts in the PER-377 cell line

A. Schematic representation of *CTGF* gene structure and probes used for northern blot analysis. Central and 3' probes (red bars) hybridise exons 3 to 5 or the 3' UTR. The 3' probe matches the Affymetrix *CTGF* probeset 209101_at target sequence (blue bar). **B.** Northern blots showing PER-377 RNA extracted at timepoints 24, 48 and 72 hrs. Blots were hybridised with either the central (left) or 3' probe (right). Canonical *CTGF* mRNA (2.4 kb) is indicated by the blue arrow. Red arrowheads indicate novel transcripts that were hybridised by both probes. A probe specific for *ATP53G* was used as a loading control. **C.** Relative abundance of canonical *CTGF* transcript at each timepoint calculated by densitometry using the signal for *ATP53G*.

5.2.1 Northern analysis of *CTGF* mRNA in ALL cell lines

While *CTGF* expression was much higher in PER-377 compared to the remaining cell lines utilised in this study, all cell lines were examined by northern blot to determine if the minor *CTGF* transcripts were expressed more generally in B-lineage ALL. Characterisation of *CTGF* mRNA was therefore performed on the B-lineage ALL cell lines described in Chapter 4, as well as the T-ALL cell line JURKAT which was employed as a negative control. Two independent northern blots utilising the two different *CTGF* probes were performed, as shown in Figure 5.2. The *CTGF*^{neg} PER-485 cell line was not available for analysis due to persistent RNA degradation occurring in three independent extractions. In agreement with the qRT-PCR data presented in Chapter 4, PER-377 displayed exceptionally high levels of *CTGF* mRNA compared to the other cell lines. For this reason, PER-377 is not shown in Figure 5.2A, due to high signal intensity with the central probe, but is shown hybridised by the 3' probe in Figure 5.2B.

Full-length *CTGF* was detectable in B-lineage ALL cell lines PER-145, PER-278 and PER-371 but not in PER-490, PER-495 or the T-ALL cell line JURKAT. These findings were in good agreement with the previous qRT-PCR experiments. Again, the central probe hybridised to 2.4 kb *CTGF* mRNA, as well as a shorter transcript of approximately 1.6 kb in PER-145, PER-278 and PER-371. The minor 1.3 kb transcript was also present in these cells lines, but at very low levels. When samples were examined with the 3' probe, the shorter transcripts observed in Figure 5.1, were evident only in the PER-377 cell line. This may be for technical reasons as the signal intensity of the 3' probe hybridised to the canonical transcript was less intense compared to the central probe in two independent experiments. Nonetheless, this experiment confirmed that the remaining *CTGF*^{pos} cell lines produce alternative *CTGF* transcripts and the features of these transcripts remained to be elucidated.

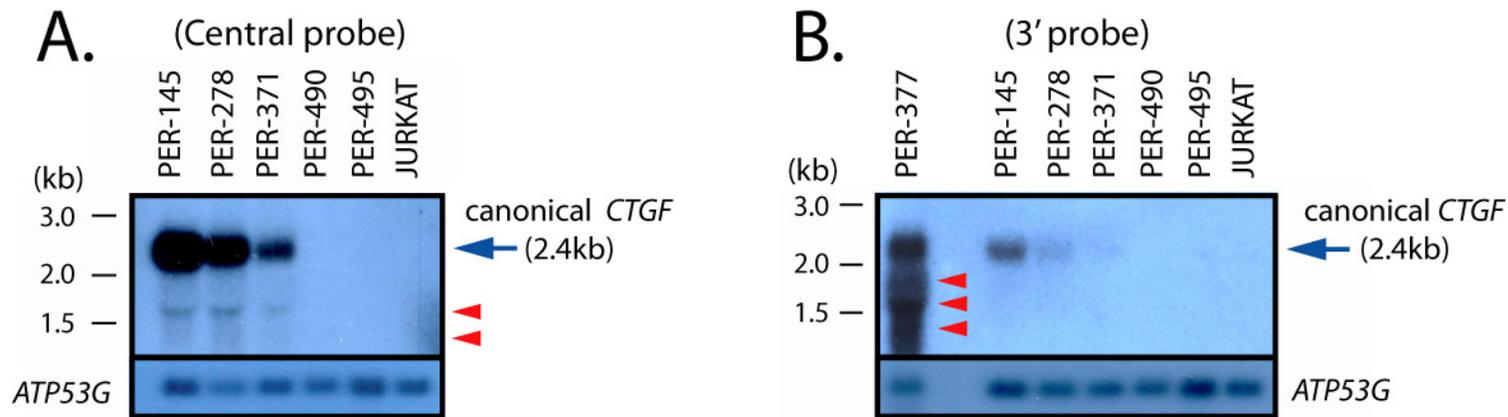


Figure 5.2 Characterisation of *CTGF* mRNA transcripts in ALL cell lines

A. Northern blot of ALL cell lines using the central probe. PER-377 was excluded from this figure due to high signal intensity. **B.** Northern blot of ALL cell lines using the 3' probe. For both blots, the canonical 2.4kb transcript is indicated by the blue arrow. Novel transcripts are indicated by red arrowheads. The T-ALL cell line JURKAT was included as a negative control. A probe designed to hybridise *ATP53G* mRNA was used as a loading control.

5.2.2 Amplification and sequencing of novel *CTGF* transcripts

Novel *CTGF* mRNA transcripts were identified by northern blotting in cell lines that were previously shown to be positive for *CTGF* expression by qRT-PCR. To examine whether these *CTGF* variants were produced by alternative splicing, a nested reverse-transcriptase PCR (RT-PCR) strategy was employed to enable their amplification and sequencing. OligodT primed cDNA was generated from *DNaseI*-treated, total cellular RNA extracts isolated from PER-377 cells in log phase growth, as they displayed the highest level of *CTGF* expression and were shown to contain these novel *CTGF* transcripts. *CTGF* cDNA was amplified from exon 1 to exon 5 and these PCR products were gel extracted, TA cloned and sequenced. Primer placement relative to the *CTGF* gene is shown in Figure 5.3A.

Somewhat surprisingly, amplification of a product corresponding to the canonical mRNA with these primers (1023 bp) was not observed. Instead, two shorter products of approximately 330 bp and 250 bp were amplified (Figure 5.3B). These PCR products were cloned and sequenced, revealing that only the larger PCR product was amplified from *CTGF* cDNA. This 331 bp product was amplified from a novel *CTGF* transcript generated by alternative splicing. Sequencing confirmed that this transcript was missing exons 2 and 3, and contained truncated forms of exons 1 and 4. Splicing of truncated exon 1 to exon 4 occurred 73 bases before the start of the *CTGF* coding region at position +207. Splicing from exon 4 to 5 adhered to canonical exon junctions. This novel cDNA was designated as Variant 1.

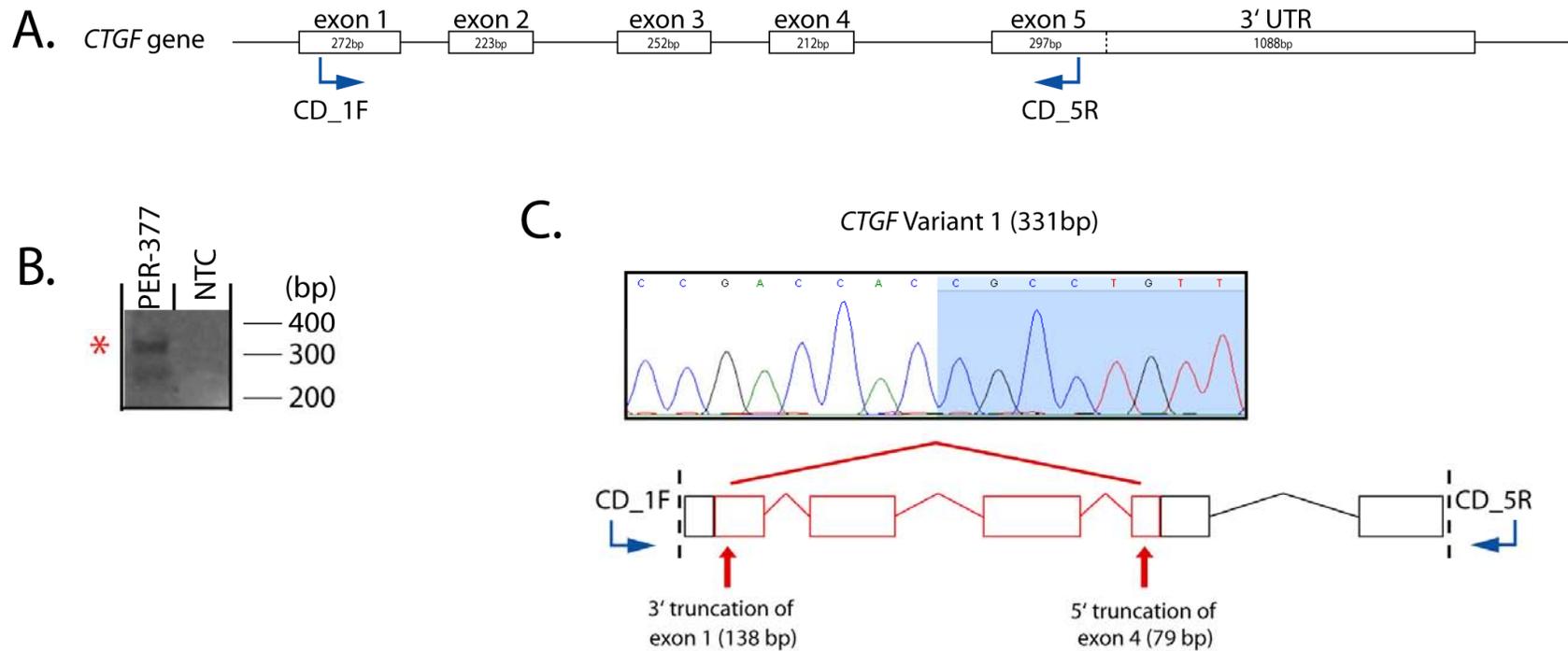


Figure 5.3 Structure of *CTGF* Variant 1

A. Schematic representation of the *CTGF* gene structure and location targeted by the CD_1F and CD_5R PCR primers. **B.** Amplification products generated from PER-377 cDNA using the CD_1F and CD_5R primers. The top product of 331bp indicated by a red asterisk was demonstrated to be a novel *CTGF* transcript (designated Variant 1) by sequencing. **C.** Sequencing the 331bp clone revealed skipping of exons 2 and 3 (shown in red) as well as truncation of exons 1 and 4.

Further extension of this PCR strategy was undertaken to capture other alternative splicing events affecting *CTGF* mRNA in PER-377 cells. A more sensitive semi-nested PCR approach was conducted by incorporating a primary round of amplification using a reverse primer within the 3' UTR of the *CTGF* transcript (CD_7R) shown in Figure 5.4A. This primary PCR was followed by a secondary, semi-nested PCR reaction using the CD_1F to CD_5R primer combination from the previous experiment. The results of this approach shown in Figure 5.4B, yielded a further two fragments of around 380bp and 400bp in addition to the previously identified Variant 1 (331bp). These larger PCR products were gel extracted and cloned, and sequencing these clones revealed that they too were *CTGF* transcript variants arising from alternative splicing events in addition to Variant 1.

These two new variants were found to be 366 bp (Variant 2) and 415 bp (Variant 3) respectively. Variant 2 exhibited splicing from a truncated exon 1, 35 bp downstream from the alternative splice site used in Variant 1. This shorter exon 1 was spliced to the same alternative acceptor splice site in exon 4 that was used to generate Variant 1, and splicing of exon 4 to exon 5 adhered to canonical exon junctions. The largest of these transcripts, Variant 3 was missing 170 bp of the 3' end of exon 1, and in contrast to Variants 1 and 2, this truncated exon 1 was spliced to a shortened form of exon 3. As with the other two variants, downstream splicing adhered to canonical exon junctions for this transcript. In all cases (Variants 1, 2 and 3) splicing of the alternative exon 1 donor splice site occurred prior to start of the canonical *CTGF* reading frame at position 207 of exon 1, thus the *CTGF* reading frame is 5' truncated in these alternative transcripts.

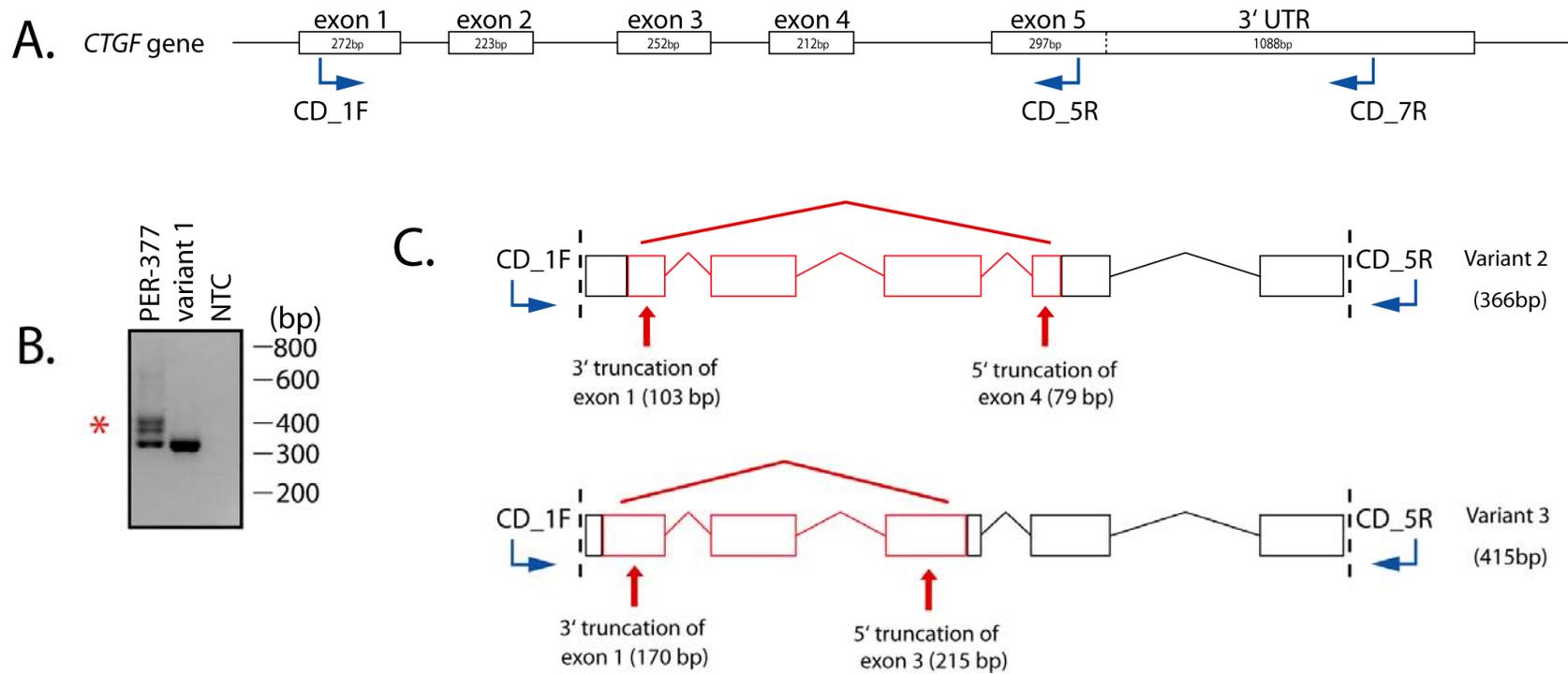


Figure 5.4 Structure of *CTGF* Variants 2 and 3

A. Schematic representation of *CTGF* gene structure and binding sites for the CD_1F, CD_5R and CD_7R PCR primers. **B.** Second round semi-nested amplification from PER-377 cDNA using CD_1F, CD_5R primers. Plasmid DNA containing the previously detected Variant 1 clone was used as a positive control in the secondary PCR. The top two bands denoted by a red asterisk were demonstrated to be *CTGF* transcript variants by sequencing. **C.** Schematic representation of alternative splicing of Variant 2 and Variant 3 *CTGF* cDNA.

5.2.2.1 Structural analysis of novel *CTGF* transcript variants

The structure of the three novel *CTGF* transcript variants outside of the region amplified by PCR was not known. However, if these three variant cDNAs retained the canonical transcriptional start and polyadenylation sites then their respective sizes would be; 1.67 kb (Variant 1), 1.7 kb (Variant 2) and 1.75 kb (Variant 3), thus any one, or all of these transcripts could represent the upper, non-canonical *CTGF* bands of approximately 1.6 kb and 1.8 kb detected in the northern blots shown in Figures 5.1 and 5.2. Splicing of all three of the variant transcripts occurred 5' of the canonical start codon. However, *in silico* analysis predicted that these transcripts retain the 3' of the canonical *CTGF* reading frame at position 102 of exon 4 (Met215). Figure 5.5 outlines the sequenced regions of Variants 1 to 3 and the position of Met215 is shown in exon 4. There is a partial match with a Kozac consensus sequence (De Angioletti et al. 2004) (gccrccAUGG) adjacent to Met215, as there is at the canonical *CTGF* start codon, thus the Met215 codon may still serve as a valid site for translation initiation.

If translated, these three transcripts could all encode a truncated *CTGF* protein containing half of the TSP1 domain and all of the CT domain. Immunoblotting experiments performed in the previous chapter with the C-terminal Rabbit-A antibody only detected the full length *CTGF* protein (38kDa) in enriched supernatants from PER-377, thus these transcripts do not appear to be translated into protein. However, this may be accounted for by technical reasons as the level of this protein isoform may be below the assay detection limit. Alternatively, the predicted shorter *CTGF* protein may not expose the epitope targeted by the Rabbit-A antibody.

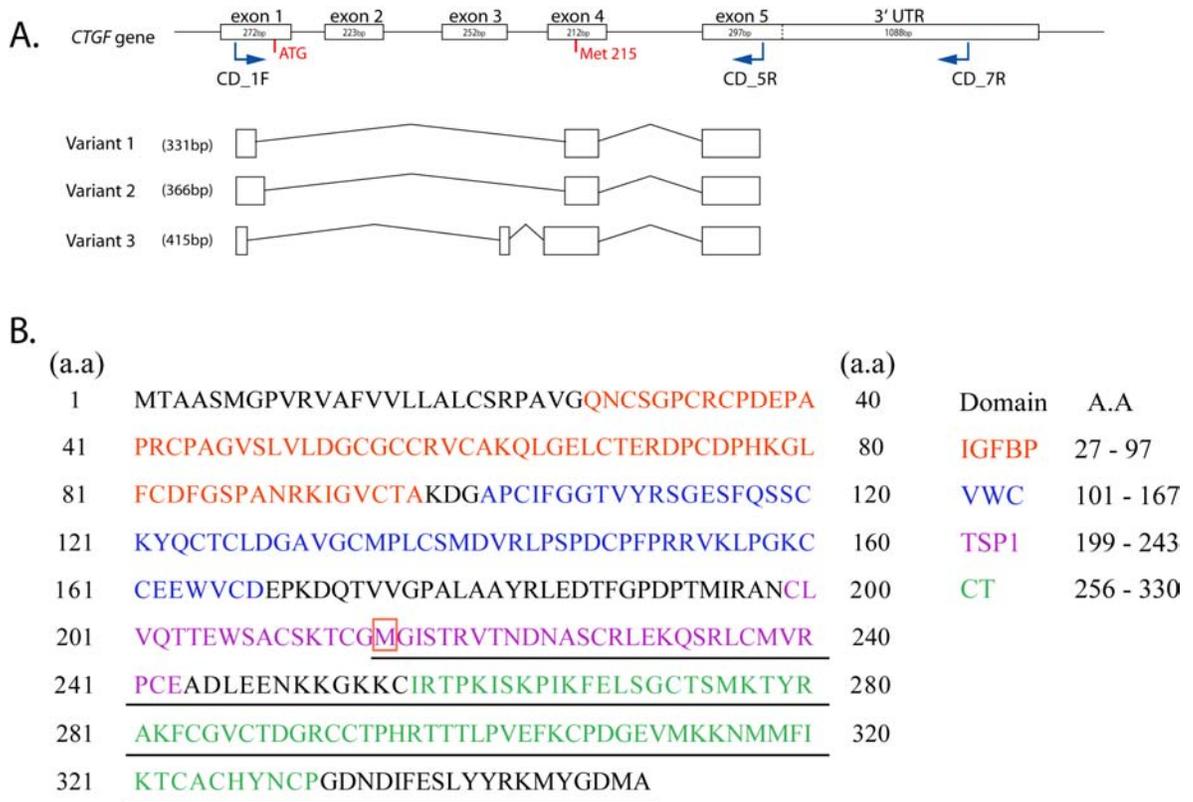


Figure 5.5 *CTGF* transcript variants may retain protein coding potential

A. *CTGF* gene structure with forward and reverse primers used to amplify Variants 1 to 3 indicated by blue arrows. A start codon (ATG) is encoded at position 207 of exon 1. Met215 encoded at position 102 of exon 4 is contained within all three variant cDNAs.

B. *CTGF* amino acid sequence. Functional domains are colour coded: IGFBP domain (red), VWC domain (blue), TSP1 domain (purple) and the CT domain (green). Translation initiation at Met215 (highlighted by the red box) can encode a 134 amino acid protein (underlined) corresponding to the C-terminal one-third of the *CTGF* protein.

5.2.3 Amplification of *CTGF* transcripts by RACE

RT-PCR amplification between exons 1 and 5 of *CTGF* from PER-377 RNA identified three novel cDNA fragments that were confirmed as *CTGF* splice variants by sequencing. It is possible that additional alternative transcripts are present in PER-377 other than those detected by the PCR approach. For this reason, 5' and 3' RACE (rapid amplification of cDNA ends) was performed in an attempt to detect the full repertoire of *CTGF* mRNA species and also to characterise the 5' and 3' ends of these variant transcripts. For both 5' and 3' RACE, two gene specific primers (GSPs) were used in nested RACE PCR reactions as outlined in Figure 5.6A. OligodT-primed first strand cDNA was generated from *DNaseI*-treated total RNA from PER-377 cells and served as a template for RACE PCR. Both 5' and 3' RACE PCR amplification from PER-377 cDNA yielded multiple RACE PCR products (Figure 5.6B). In the case of 5' RACE, the full length canonical transcript (1064bp) was not amplified in either primary or secondary PCRs, however several smaller products between ~700 bp and 200 bp were amplified. When 3' RACE was performed, there was strong amplification of a product at the expected size of the canonical *CTGF* transcript (2.4 kb) together with products at approximately 2.8 kb and 1 kb and less abundant products at 500 bp and 200 bp.

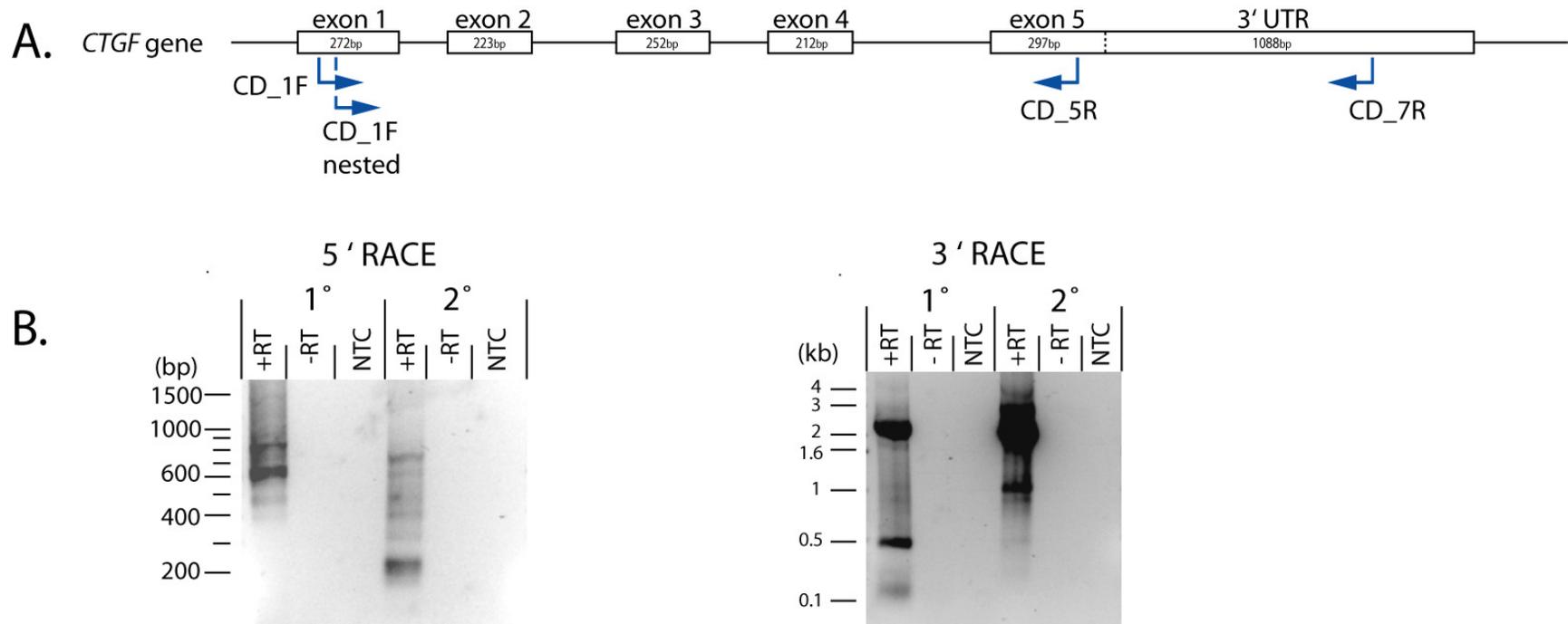


Figure 5.6 Amplification of *CTGF* variants by 5' and 3' RACE.

A. Location of gene specific primers (GSPs) used for RACE amplification. CD_1F and CD_1F_nested were used as GSPs for 3' RACE, while CD_7R and CD_5R were used as GSPs for 5' RACE. **B.** Amplification of PER-377 cDNA by 5' RACE and 3' RACE primary (1°) and secondary (2°) PCRs. Reverse transcriptase negative (-RT) controls from 1° PCR reactions were used as templates for 2° PCR (-RT) controls. No template controls (NTCs) were used for both reactions.

5.2.3.1 Isolation of RACE products by band-stab PCR

The distribution of RACE products shown in Figure 5.6B indicated that isolation of individual products would be difficult. In the case of 3' RACE, there was a clear amplification bias for the more abundant canonical transcript, which would result in an overrepresentation of this transcript if shotgun cloning was used to clone unselected RACE PCR products. In the case of 5' RACE, there were a large number of bands that did not resolve satisfactorily, even on high percentage agarose gels. For this reason, 5' and 3' RACE products of interest were isolated by using a band-stab PCR strategy (Bjourson & Cooper 1992). The band-stab PCR allowed the selective isolation of individual RACE products that would otherwise have been difficult to independently obtain as shown in Figure 5.7. For the 5' RACE PCR (Figure 5.7A), six bands were identified as independent PCR products and were targeted for re-amplification by band-stab PCR, using the GSP primer CD_5R and the secondary 5' RACE primer. For the 3' RACE PCR (Figure 5.7B), five bands were similarly selected and then re-amplified using the CD_1F_nested primer and the nested 3' RACE primer.

When 5' RACE products were re-amplified by band-stab PCR, the resulting products displayed the same size as those bands used as the template or target band in band-stab reactions (Figure 5.7A, lane 4, bands 1 to 6). Separation of these PCR products provided superior resolution compared to RACE PCR alone, facilitating the independent isolation of target bands by gel extraction. Separation of band-stab PCR products from 3' RACE revealed there was amplification of a product in lane 1 and 2 that corresponded to the size of the canonical *CTGF* transcript. Lane 3 contained a dominant product of approximately 1.5 kb, while lanes 4 and 5 contained dominant products of 600 and 250 bp respectively.

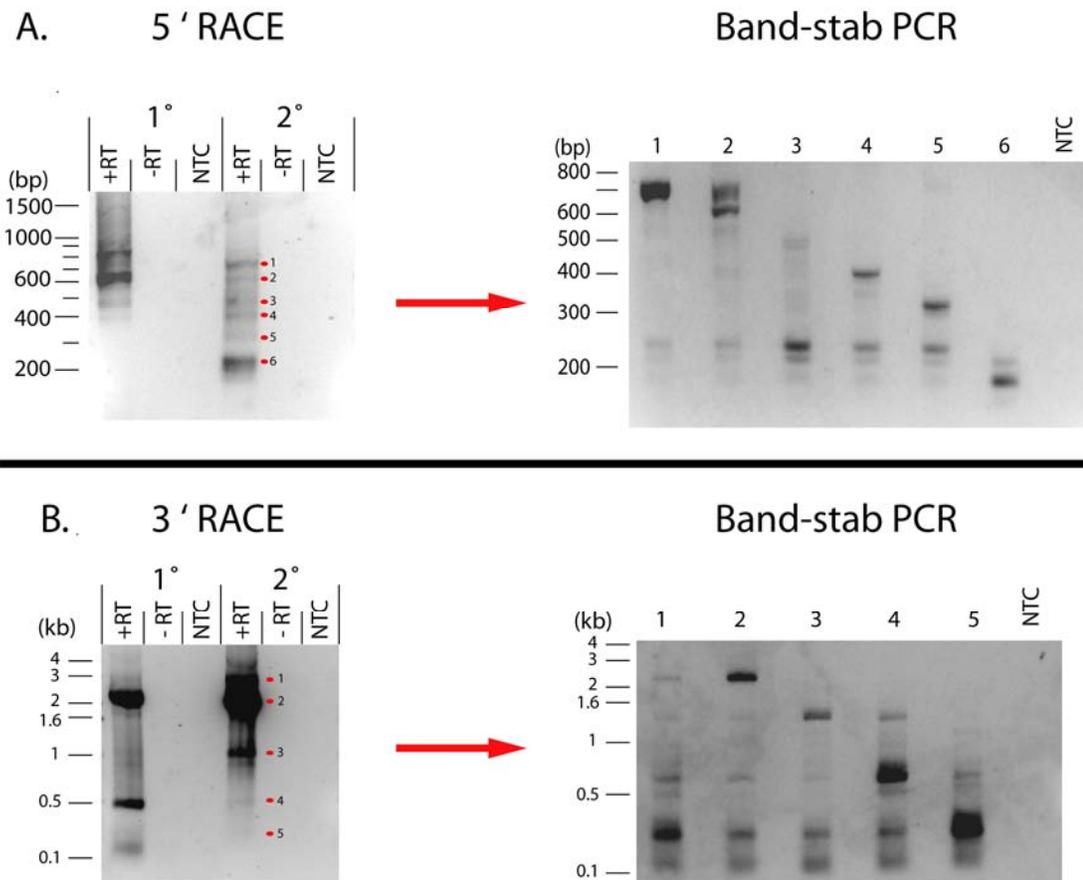


Figure 5.7 Re-amplification of individual *CTGF* RACE products by band-stab PCR

(A) 5' RACE products generated from PER-377 cDNA. Individual target bands 1 to 6 (red dots; left panel) from the nested secondary RACE PCR were re-amplified by band-stab PCR using the CD_5R GSP and the nested 5' RACE primer (right panel). (B) 3' RACE product generated from PER-377 cDNA. Individual target bands 1 to 5 (red dots; left panel) were re-amplified by band-stab PCR using the CD_1F_nested GSP and the nested 3' RACE primer (right panel).

5.2.3.2 Validation of a *CTGF* probe for Southern blot hybridisation

Before extracting and cloning RACE band-stab cDNAs, Southern hybridisation was used to verify that these products contained genuine *CTGF* sequences. Only those bands

hybridised by a radio-labelled *CTGF* specific probe were isolated for further analysis. The *CTGF* probe was generated by amplifying from exon 1 to 5 from a cloned, full length *CTGF* cDNA, using the CD_1F and CD_5R primers. The specificity of the *CTGF* probe was examined by hybridisation to unselected primary and secondary *CTGF* 5' RACE PCR products which displayed multiple bands by gel electrophoresis. As shown in Figure 5.8B, the *CTGF* probe strongly hybridised cDNA from the secondary RACE PCR only, indicating that the probe was highly specific for *CTGF* and was therefore suitable for probing band-stab PCR products.

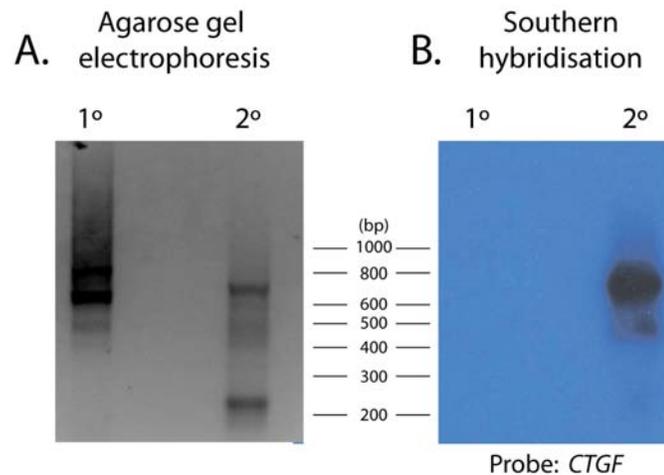


Figure 5.8 Southern blotting of 5' RACE PCR products

A. Primary (1°) and secondary (2°) PCR products from PER-377 5' RACE, separated by agarose gel electrophoresis and stained with ethidium bromide. **B.** Southern hybridisation of these samples, probed using a radio-labelled (³²P) *CTGF* probe.

5.2.3.3 Characterisation of 5' RACE products

Southern hybridisation of 5' RACE band-stab PCR products was performed with the *CTGF* probe to identify genuine *CTGF* RACE PCR products. As shown in Figure 5.9A, individual bands that were clearly hybridised by the *CTGF* probe (marked by red boxes) were deemed to be genuine *CTGF* RACE PCR products and were isolated and cloned.

Lane 2 in Figure 5.9A contains two prominent bands between 600 and 700 bp in size. The upper band was similar in size to the predominant product in lane 1 while the lower band did not appear to specifically hybridise the *CTGF* probe and as such, lane 2 products were not included in further analysis. A total of seven bands were isolated and cloned and in the case of lanes 3 and 6, two individual bands were isolated from each lane. Three independent clones from each transformation were isolated and sequenced. Sequencing of these twenty one 5' RACE clones revealed a total of eleven distinct *CTGF* transcripts, as shown in Figure 5.9B. Overall, there was a high degree of heterogeneity observed in the 5' end of 5' RACE clones and this heterogeneity suggests that there are likely to be additional *CTGF* transcripts present that were not fully characterised in these experiments. Of note, splicing of downstream intron/exon junctions were in agreement with canonical mRNA exon boundaries. Those 5' ends within exonic regions are colour coded in red, while transcripts displaying 5' ends localised within introns were similarly colour coded, however these were grouped according to their intron of origin, yellow for intron 2, blue for intron 3, and green for intron 4.

Sequencing these RACE products revealed that clones 5_1:c1-c3 were between 688 and 706 bp in size and exhibited 5' ends ranging from 89 to 108 bp upstream from the exon 3 boundary. Clones 5_3A:c1-c3 were isolated from the upper band in lane 3 of Figure 5.9A. All three clones were identical in size (476 bp) and structure, exhibiting a 5' end contained within intron 3, 129 bp upstream from the canonical exon 4 boundary and only 3bp downstream from the canonical exon 3 donor splice site. The 5_3B:c1-c3 clones derived from isolates of the lower band in lane 3 of Figure 5.9A were also identical in size and structure. These cDNAs were 222 bp in length and exhibited a 5' start site 89 bp upstream from the canonical end of exon 4.

Clones derived from lane 4 in Figure 5.9A were similar in size but differed in their 5' ends and structure. Clones 1 and 2 (5_4:c1,c2) both exhibited 5' ends originating within exon 3, just before the canonical exon 3 donor site boundary, while the 5' end of clone 3 (5_4:c3) originated from within intron 3, 48 bp upstream from the canonical exon 4

acceptor site. The three 5_5:c1-c3 clones were 326 bp in size and exhibited a 5' end originating within intron 4, 191 bp upstream from the 5' end of exon 5. The upper isolate from lane 6 in Figure 5.9A (55_6A) was 197 bp in length and contained 62 bp of the 3' end of exon 4 spliced to exon 5. All three clones were identical. Clones derived from 5_6B were also identical and contained 36 bp of exon 4 spliced to exon 5. None of these 5' RACE clones contained any part of exons 1 or 2 in their 5' end. This may be for technical reasons since, as stated previously, the canonical *CTGF* cDNA could not be obtained by 5' RACE and exons 1 and 2 have a high GC content. Similarly, the cDNA variants detected by RT-PCR in the previous section were not detected by 5' RACE.

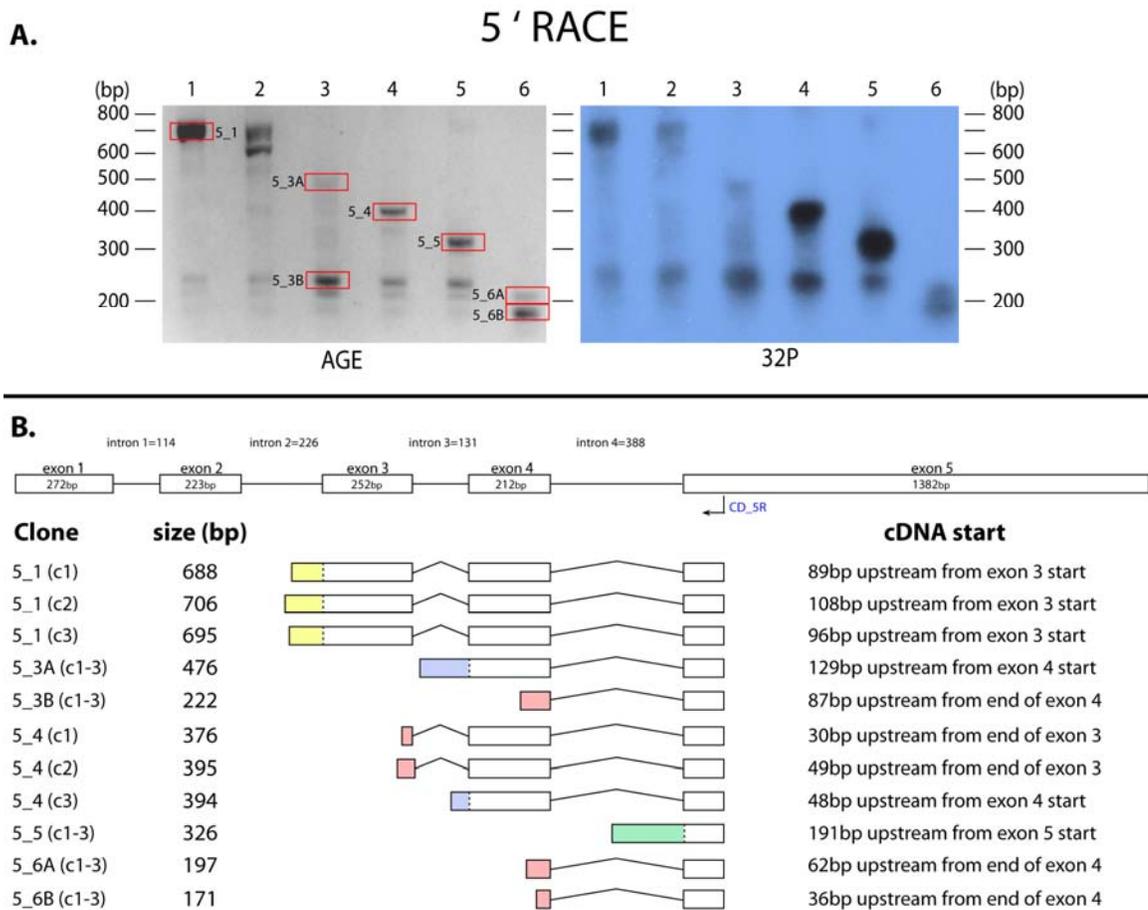


Figure 5.9 Characterisation of PER-377 *CTGF* 5' RACE products

A. Band-stab isolated RACE products separated by agarose gel electrophoresis (AGE) and then examined by Southern blotting with a radio-labelled (^{32}P) *CTGF* probe. Red boxes indicate bands that were isolated, cloned and sequenced. Where two or more isolates were obtained from a single lane, isolates are labelled by adding A or B to name.

B. Schematic representation of 5' RACE clones aligned with the *CTGF* gene. Reverse primer CD_5R is shown in blue. Clone name is followed by identifier for sequenced clone (c1, 2 or 3), and the size of clones is shown in bp. The cDNA start site is indicated on the right of each RACE clone. Vertical broken lines in transcripts represent canonical exon junctions. Clones with 5' ends originating within canonical *CTGF* exons have their first exon coloured in pink, while those displaying 5' ends within intronic regions have the intronic portion of their 5' end colour coded as yellow for intron 2, blue for intron 3 and green for intron 4.

5.2.3.4 Characterisation of 3' RACE products

Southern hybridisation of 3' RACE PCR products re-amplified by band-stab PCR (Figure 5.10A) was performed as previously described for 5' RACE PCR products. Hybridisation with the *CTGF* specific probe revealed that lane 1 did not contain any *CTGF* positive products that were unique compared with other lanes and as such, this lane was excluded from further analysis. A single band was extracted from each of the remaining 4 band-stab PCR reactions (lanes 2 to 5) as indicated by the red boxes in Figure 5.10A. Sequencing of the cloned isolates was performed with the primers CD_4F, CD_5F, CD_6R and CD_7R which anneal within exonic regions of the *CTGF* gene as indicated in Figure 5.10A. The 3' end of cloned band-stab isolates were sequenced using vector specific primers (M13 forward and reverse).

Sequencing of the 2.3 kb clones derived from lane 2 in Figure 5.10A (3_2:c1-c3) revealed that this 3' RACE PCR product was amplified from the full length, canonical *CTGF* transcript. This is in contrast to 5' RACE which did not yield products amplified from the canonical *CTGF* transcript. Analysis of 3_3:c1-c3 clones revealed that this isolate was amplified from a novel *CTGF* transcript that exhibited premature polyadenylation within exon 5. This transcript contained 384 bp of exon 5 and harboured an alternative polyadenylation site located 87 bp downstream from the canonical translation termination codon. Thus, this transcript retained *CTGF* coding potential, but was devoid of established regulatory elements contained within the 3' UTR (Ohgawara et al. 2009; Duisters et al. 2009; Ernst et al. 2010; Kubota et al. 2005). If transcribed from the canonical transcription start site, this mRNA variant would be 1.36 kb in length, thus this transcript may represent the lower band detected by northern blotting of B-lineage ALL cell lines, see Figures 5.1 and 5.2. This mRNA variant may represent a long-lived transcript with increased translational potential due to the lack of the 3' regulatory elements. Thus, while this minor transcript was only present at low concentrations, it may nevertheless have biological importance.

Sequencing of 3_4:c1-c3 demonstrated that these isolates were cloned from a PCR artefact generated through reverse priming by the CD_1F nested primer, and not a genuine RACE product. Reverse priming occurred despite low specificity between primer and target, with only five bases of complementarity wholly localised to the 3' end of the primer. Sequencing of the final RACE PCR isolate (3_5:c1-c3) revealed a novel RACE product that contained 121bp of *CTGF* exon 1, followed by a 46 bp of unknown sequence and a putative polyadenylation signal (shown in more detail in Figure 5.11A). Using the BLAST sequence alignment tool, this unknown sequence was used to query the human RNA_refseq database for a potential sequence match. Only one mRNA transcript; leucine rich repeat containing 58 (*LRRC58*) mRNA, matched the sequence query with 100 percent identity, shown in Figure 5.11B. Furthermore, when the 3_5 clone sequences were aligned with the *LRRC58* genomic sequence (Figure 5.11C), it was apparent that the *LRRC58* genomic sequence contained a run of adenines similar to those seen in the sequenced RACE clone indicating that this segment of the sequenced clones may not represent the 3' polyadenylation site of the original mRNA, thus the full length mRNA transcript may have been considerably longer.

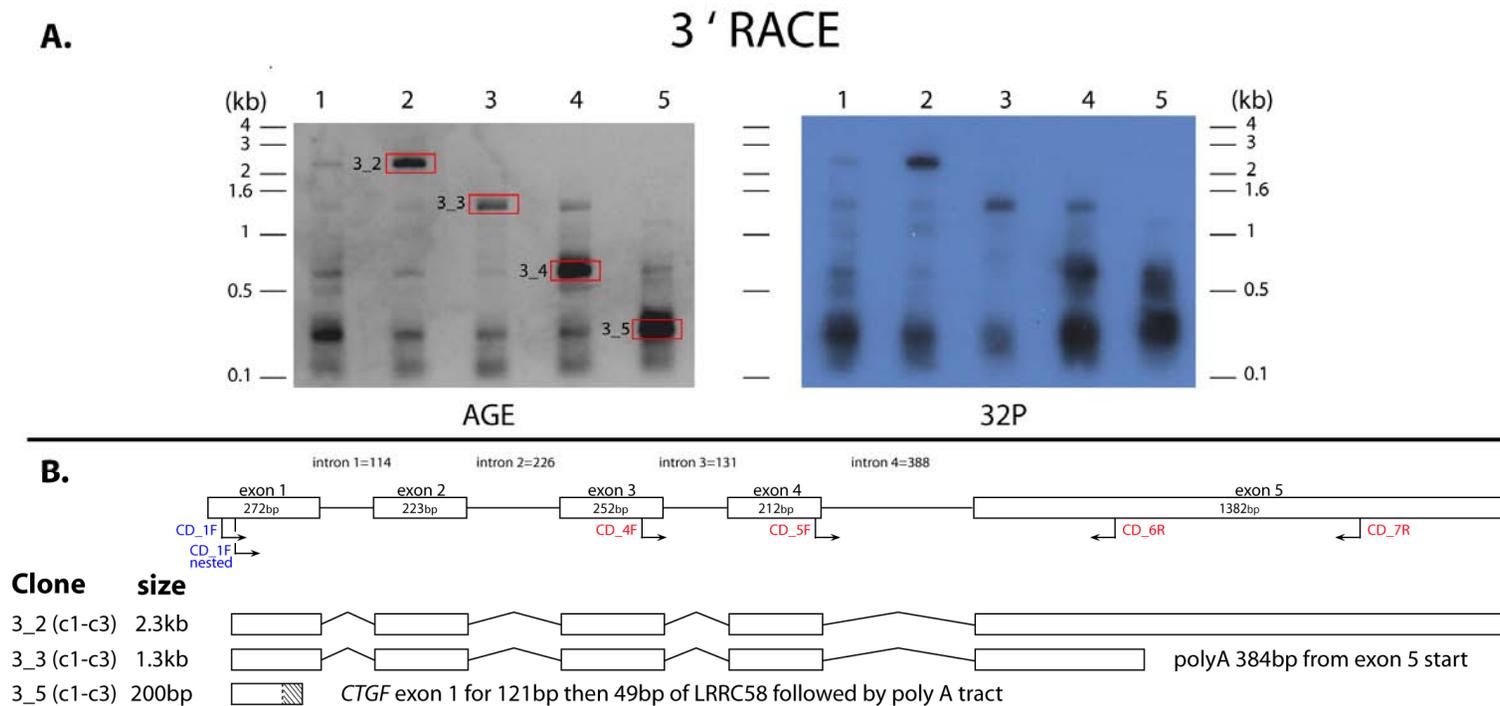


Figure 5.10 Characterisation of PER-377 *CTGF* 3' RACE products

A. Band-stab isolated RACE products were separated by agarose gel electrophoresis (AGE) and then examined by Southern blotting with a radio-labelled (^{32}P) *CTGF* probe. Red boxes indicate bands that were isolated, cloned and sequenced. **B.** Schematic representation of 3' RACE clones aligned with the *CTGF* gene. Clone ID and size are indicated on the left. Forward GSPs for RACE PCR shown in blue. Primers used for internal sequencing of clones shown in red. M13 forward and reverse vector primers were used to sequence insert ends.

These observations suggested that there may be a translocation or insertion event involving the *LRR58* and *CTGF* gene loci. This seemed unlikely given the results reported Chapter 4 that ruled out rearrangement of the *CTGF* locus in PER-377, however to test this possibility two reverse primers targeting the *LRR58* locus immediately upstream from the run of adenines in Figure 5.11C were used in conjunction with the CD_1F and CD_1F_nested PCR primers. It was hypothesised that if the 3_5 RACE clone arose from a fusion between these two genomic loci these primers would facilitate amplification of the fusion gene. Amplification with these primer combinations using PER-377 genomic DNA did not yield any products, ruling out the existence of such a fusion in this cell line. Therefore this fusion transcript most likely arose as a result of aberrant trans-splicing between two disparate pre-mRNA transcripts and is unlikely to represent a genuine coding transcript (Houseley & Tollervey 2009; Li et al. 2009b).

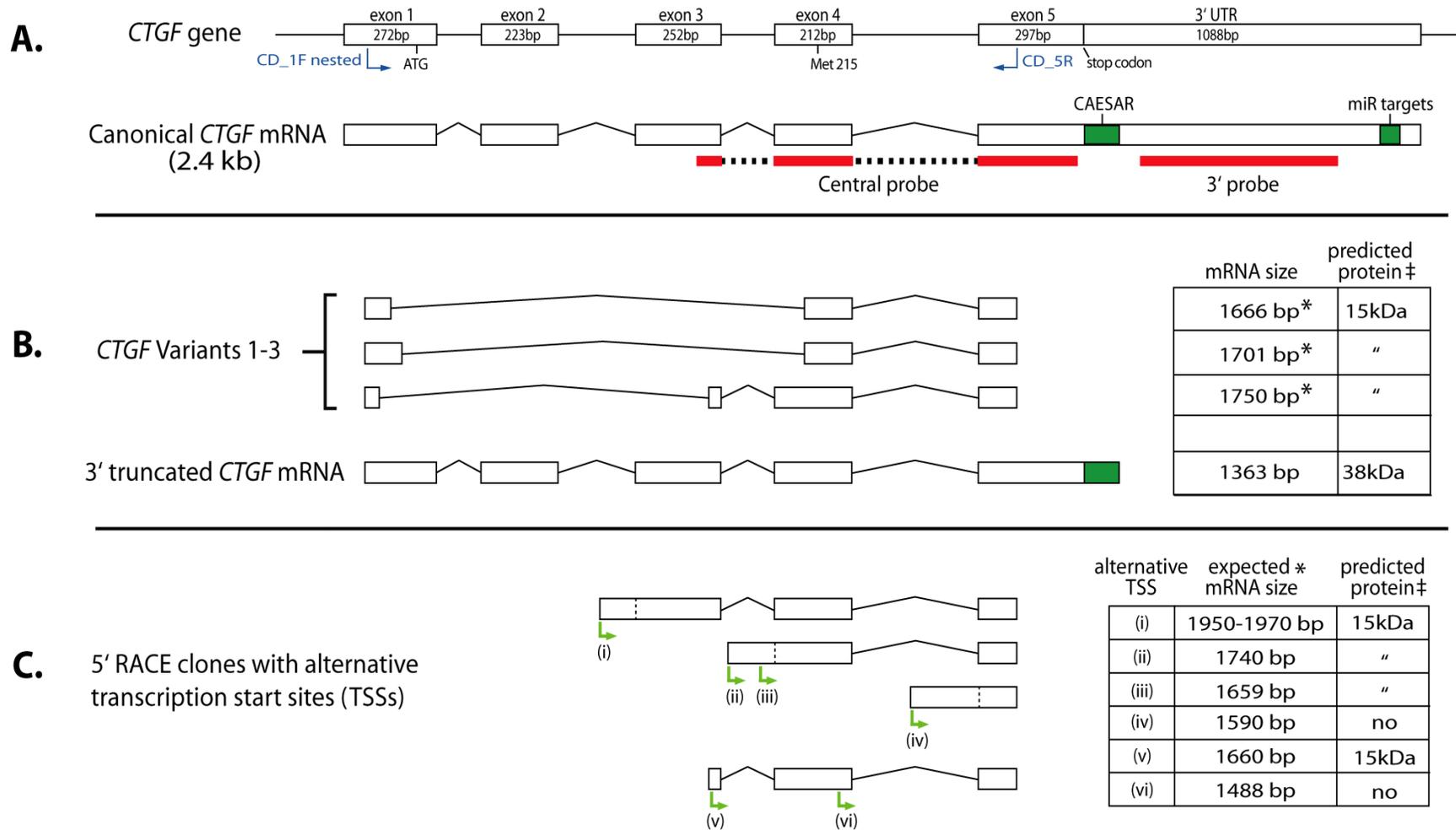


Figure 5.12 Summary of novel *CTGF* transcripts detected in pre-B ALL cells (Figure legend over page)

Figure 5.12 Summary of novel *CTGF* transcripts detected in pre-B ALL cells

A. Schematic representation of canonical *CTGF* mRNA aligned with the *CTGF* gene. Forward and reverse primers used in RACE experiments are shown in blue. The Met 215 codon which may serve as a secondary translation start site, is shown in exon 4. The Central and 3' northern blot probes (red bars) are aligned to the *CTGF* mRNA transcript. Post-transcriptional regulatory elements contained within the *CTGF* 3' UTR are shown in green. **B.** Novel *CTGF* transcripts detected by RT-PCR (Variants 1 to 3) are shown together with the 3' truncated RACE transcript lacking most of the 3' UTR. For each of these transcripts, the mRNA size is shown. *Predicted assuming canonical TSS and/or polyadenylation sites were present outside of sequenced regions. The predicted protein coding potential of transcripts is also shown ‡*In silico* predictions of protein coding potential assumed the presence of an intact exon 5 including canonical stop codon. **C.** Representation of 5' RACE clones identified in this study that exhibited non-canonical 5' ends. Vertical broken lines represent canonical exon boundaries. Transcripts are grouped by the location of their 5' ends, either intron2, intron 3, intron 4 or those transcripts that originated within exonic regions. For each of the four groups, alternative transcription start sites (alt. TSS) are shown by numbered green arrows (i to vi) and annotated on the right, with expected mRNA size* and predicted protein coding potential‡.

5.3 Discussion

The disparity between *in vitro* *CTGF* mRNA and protein levels documented in Chapter 4 prompted a more detailed analysis of *CTGF* mRNA to confirm that those *CTGF* transcripts detected by qRT-PCR and microarray methods represented the canonical 2.4 kb message. The present findings demonstrate that the canonical *CTGF* mRNA was indeed the major form of *CTGF* transcribed in B-lineage ALL cell lines. Furthermore, canonical *CTGF* transcripts were readily detectable in the *CTGF*^{low} cell lines PER-145, PER-278 and PER-371 by northern blotting. This confirmed that these *CTGF*^{low} cell lines generate *CTGF* mRNA, despite lacking detectable levels of CTGF protein, as documented in Chapter 4.

In addition to canonical *CTGF* mRNA, northern blotting revealed the presence of shorter, much less abundant *CTGF* transcripts produced in *CTGF* positive cell lines. Sequencing of these mRNA variants (isolated from the *CTGF*^{high} cell line PER-377), confirmed that these minor *CTGF* transcripts, estimated to be present at less than 10% of the total pool of *CTGF* mRNA, were produced by alternative pre-mRNA splicing. This is the first evidence of alternative splicing of *CTGF* mRNA in any cell type and supports further investigation into alternative *CTGF* splicing in other cancers and normal tissues. A number of alternative splicing events were observed in sequenced cDNA fragments and RACE PCR products generated from the PER-377 cell line. These non-canonical transcripts summarised in Figure 5.12, exhibited exon skipping, alternative 5' (donor) and 3' (acceptor) splice site usage, alternative or premature polyadenylation and alternative transcription start sites.

5.3.1 Aberrant *CTGF* transcription

Targeted RT-PCR amplification of *CTGF* cDNA from exon 1 to 5 revealed the presence of three novel transcripts (Figure 5.12B). These transcripts exhibited internal deletions of exons 2 and 3, alternative exon 1 donor splice sites, and alternative exon 3 and 4 acceptor splice sites. *In silico* analysis of these alternatively spliced transcripts indicated that they have the potential to encode a shorter, N-terminal truncated *CTGF* protein of approximately 15kDa. This protein would lack N-terminal domains including the signal peptide which may impair secretion of this truncated *CTGF* protein. Several smaller MW *CTGF* proteins ranging in size from 10 kDa up to 20 kDa have been detected in human cells and biological fluids (Steffen et al. 1998; Yang et al. 1998), however their existence has uniformly been attributed to post translational proteolysis of the full length *CTGF* protein. This notion may need to be reconsidered given the current findings, particularly if alternative splicing of *CTGF* is demonstrated in other cell types. The ability of these truncated *CTGF* cDNAs to encode biologically active proteins should be examined *in vitro*. If such protein isoforms are confirmed, functional characterisation will be required to establish their biological function and sub-cellular localisation.

In addition to alternative splicing events identified by RT-PCR, analysis of *CTGF* transcripts by 3' RACE uncovered a novel 1.3 kb *CTGF* variant with an early polyadenylation signal 87 bp after the end of the *CTGF* coding region (Figure 5.12B). This transcript contained an intact coding region, but did not contain the canonical 3' UTR and thus was devoid of important regulatory elements that have been shown to contribute to the post-transcriptional regulation of *CTGF* expression in other cell types. These include target sites for miRNAs, including miR18a, miR30, miR130 and the miR-17-92 cluster (Ohgawara et al. 2009; Duisters et al. 2009; Ernst et al. 2010). While these miRNA target sites were missing from this 1.3 kb transcript, the CAESAR element described by Kubota and colleagues remained intact (Kubota et al. 2005). If this transcript was initiated at the canonical TSS, which is likely given the proximity of the gene specific forward primer CD_1F to the annotated *CTGF* TSS (Figure 5.12A), then the predicted size of this transcript would be 1.362 kb. This is the approximate size of one of the non-canonical transcripts identified by northern blots and points to the likely abundance of this transcript in *CTGF* positive B-lineage ALL cell lines, albeit at a minor frequency. Increased early polyadenylation of mRNA transcripts is a general feature of cancer cells and has been shown to enhance the protein coding potential of transcripts when compared to canonical mRNA by up to ten-fold (Mayr & Bartel 2009). Recombinant expression of this cDNA may reveal a higher translational affinity compared to canonical *CTGF* owing to differences in the 3'UTR. Thus, this 1.3 kb *CTGF* transcript represents an important candidate for further investigation and efforts should be made to determine its biological distribution in ALL cells and other tissues.

Analysis of *CTGF* mRNA conducted by 5' RACE detected a range of cDNAs with different 5' ends ranging from internal regions of intron 2 to intron 4, summarised in Figure 5.12C. These non-canonical 5' ends may represent alternative TSSs, suggesting possible promoter activity within the *CTGF* coding region. Analysis of 5' RACE clones indicated that there were six regions that served as alternative TSSs (shown as green arrows in Figure 5.12C), one in intron 2, two in intron 3 one in intron 4 and on each in the 3' end of exons 3 and 4. As with the *CTGF* cDNA fragments identified by RT-PCR,

some of these novel transcripts are predicted to encode shorter CTGF protein isoforms if translation is initiated at Met 215 and the coding portion of exon 5 remains intact. A limitation of the present study is that 5' RACE clones were amplified using a GSP upstream from the canonical stop codon (CD_5R), and thus the downstream composition and coding potential of the 5' RACE clones is not certain.

A recent report suggests that the vast majority of protein coding genes are affected by alternative TSS usage, however the causes behind this phenomenon remain unclear (Denoed et al. 2007). Alternative TSS selection has not been reported to affect human *CTGF* transcription, however a recent report by Huang and colleagues demonstrated that the murine *CTGF* 5' UTR can induce reporter gene activation through promoter-like sequence elements (Huang et al. 2007). While the 5' end of the human *CTGF* gene has not been investigated for promoter-like activity, the present findings suggest that such a study may be warranted as this may account for the alternative TSSs observed in the present study.

5.3.2 Post transcriptional regulation of *CTGF* mRNA

In assessing biological specimens, and in particular large patient cohorts for expression of a gene or genes of interest, microarray technology has paved the way for large-scale, high-throughput analysis. However, despite the importance of gene expression data, recent studies have suggested that mRNA levels are only a weak surrogate for the expression of proteins they encode (Tian et al. 2004; Nie et al. 2006), capturing only around 40% of global changes in protein expression levels. In agreement with this notion, *CTGF* mRNA was not a direct predictor of whether ALL cells actively synthesise CTGF protein. This may have been for technical reasons, because while low mRNA levels can be detected by RT-PCR and northern blot, correspondingly low protein levels may be below the detection limit of immunoblotting. However, another distinct possibility is that post-transcriptional regulation impacts upon the active synthesis of CTGF protein in these *CTGF*^{low} cell lines. Post-transcriptional regulation of gene expression has emerged as of key importance in recent years (Volinia et al. 2010;

Ruan et al. 2009), and there is significant evidence that miRNAs can modulate CTGF protein expression (Ohgawara et al. 2009; Duisters et al. 2009; Ernst et al. 2010). Investigation of miRNA levels in ALL was considered beyond the scope of the present study. However, analysis of the levels and activity of *CTGF* associated miRNAs may reveal an association between their activity and CTGF protein levels in pre-B ALL and such a study should be considered in the future.

The importance of being able to accurately classify the CTGF protein status of ALL cells is highlighted by a 2007 study conducted by Sala-Torra and colleagues (Sala-Torra et al. 2007). Analysis of adult ALL specimens revealed that a poor outcome was associated with high *CTGF* expression. Of particular note in light of the current findings however, is that when patients were grouped into low, intermediate or high *CTGF* mRNA expression groups, the 5 year overall survival of patients was 58, 42 and 11% respectively (Sala-Torra et al. 2007). These data indicate that patients with high levels of *CTGF* mRNA exhibited markedly poorer survival compared to those with low or intermediate levels of *CTGF* mRNA. The average survival rate for adults diagnosed with ALL is around 40% (Rowe et al. 2005; Annino et al. 2002), thus the group with high *CTGF* mRNA had a significantly poorer outcome.

The Sala-Torra report, coupled with the findings made in the present study suggest that there may be a stochastic mRNA tipping point that must be achieved before CTGF protein synthesis can occur. It is feasible that as the level of *CTGF* mRNA increases, the ability of post-transcriptional regulatory elements to hold CTGF protein synthesis in check is curtailed. No analysis of CTGF protein expression in pre-B ALL patient specimens has been reported to date and it is clear that such a study is now warranted to correlate *CTGF* mRNA with protein levels and furthermore to ascertain if stratifying patients based upon CTGF protein expression holds more robust prognostic significance compared to mRNA levels alone.

5.3.3 Conclusions

The experiments reported in this chapter confirmed that while the major form (>90%) of *CTGF* mRNA expressed in B-lineage ALL cell lines was canonical, there was also a range of alternative *CTGF* transcripts produced in these cell lines including one transcript missing 3' regulatory elements. Furthermore, the observation that *CTGF*^{low} cell lines transcribe canonical *CTGF* suggests that post-transcriptional regulation may be limiting CTGF protein synthesis in these cell lines. These findings are the first sequence-level evidence of alternative splicing of *CTGF* mRNA and should prompt future research aimed at elucidating the full extent of alternative splicing affecting *CTGF* expression in both healthy and diseased tissues.

Chapter 6

Epigenetic Mechanisms

Influencing *CTGF* Expression in

Pre-B ALL

Chapter 6

Epigenetic Mechanisms Influencing *CTGF* Expression in Pre-B ALL

6.1 Introduction

Epigenetic changes are now accepted as a hallmark of many cancers, including those of hematopoietic origins (Bonifer & Bowen 2010; Martin-Subero et al. 2009). Cytosine methylation represents a reversible, epigenetic change that occurs predominantly at CpG dinucleotides in the mammalian genome, and was first suggested to play a role in gene expression and differentiation in the mid 1970's (Holliday & Pugh 1975; Riggs 1975). The notion that methylation of cytosine residues can result in altered gene expression without changing the underlying DNA sequence is now widely accepted, but still poorly understood (Watanabe & Maekawa 2010). This is evidenced by the recent discovery of 5'-hydroxy-methylcytosine (OH-m⁵C) which is indistinguishable from 5'-methylcytosine (m⁵C) by traditional bisulfite conversion-based analysis and is postulated to have functions distinct from that of m⁵C (Kriaucionis & Heintz 2009; Nestor et al. 2010). As well as covalent modifications to DNA, post-translational modification of histones, the fundamental building blocks of the nucleosome, can alter their interaction with DNA and other nuclear proteins, impacting significantly upon the transcriptional potential of surrounding DNA. A large number of histone modifications with the ability to alter gene expression have been characterised including acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ADP ribosylation, deimination and proline isomerisation (Kouzarides 2007). Together, DNA methylation and histone modifications represent an epigenetic code with enormous biological importance.

Genomic hypomethylation was the initial epigenetic abnormality identified in human cancers almost 30 years ago (Feinberg & Vogelstein 1983a; Gama-Sosa et al. 1983). These seminal reports identified a significant decrease in the overall m⁵C content in primary cancer cells compared to healthy cells isolated from adjacent tissue. Furthermore, this reduction in m⁵C was progressive and further enriched in metastases

compared with the original tumours, suggesting that increased hypomethylation was associated with a more aggressive or metastatic phenotype (Feinberg & Vogelstein 1983a; Gama-Sosa et al. 1983). A raft of subsequent studies have demonstrated genomic hypomethylation in numerous cancer types, including prostate, bladder, cervical, and brain, as well as chronic lymphoblastic leukaemia (Brothman et al. 2005; Seifert et al. 2007; de Capoa et al. 2003; Cadieux et al. 2006; Wahlfors et al. 1992). A global reduction in DNA methylation is likely to significantly impact upon gene expression within cancer cells by promoting transcriptional activation of regions that are normally silent.

Aberrant chromatin changes are frequently observed in MLL-rearranged leukaemia and contribute to leukaemogenesis through epigenetic activation of *HOXA* and *MEIS1* genes, which promote uncontrolled self-renewal (Guenther et al. 2008; Somerville & Cleary 2010). The role of DNA methylation in leukaemogenesis is less clear, however it is generally accepted that DNA methylation can inhibit binding by trans-factors to DNA and can also recruit protein complexes that remodel chromatin and promote gene inactivation (Feng & Zhang 2001). Thus, an increase in DNA methylation at gene promoters is generally associated with inactivation of gene transcription and may contribute to leukaemogenesis by silencing of tumour suppressor loci (Vucic et al. 2008; Bernstein et al. 2007).

Studies reporting hypermethylation of classical tumour suppressor genes (TSG) have dominated the scientific literature (Zhang et al. 2007; Sidhu et al. 2005; Ebinger et al. 2004; Honorio et al. 2003), however almost all cases of TSG silencing through focal hypermethylation is accompanied by a general hypomethylation of the cancer genome, and these changes can result in activation of proto-oncogenes (Ehrlich 2009). Feinberg and colleagues provided the first evidence of this when they identified hypomethylation of *HRAS* and *KRAS* in lung and colonic adenocarcinomas compared to normal tissue (Feinberg & Vogelstein 1983b). Activation of proto-oncogenes associated with a reduction in DNA methylation in gene regulatory bodies has now been recorded in numerous cancer types, including prostate, liver, chronic lymphocytic leukaemia and

carcinoma (Wang et al. 2007; Tsujiuchi et al. 1999; Hanada et al. 1993; Smith et al. 2009).

Experiments reported in the previous chapters did not identify mutations or genetic changes that could account for the aberrant *CTGF* expression in pre-B ALL. Therefore, it may be possible that changes in *CTGF* gene expression arise due to altered CpG methylation at the *CTGF* locus. As discussed in Chapter 1, there is some evidence for the involvement of epigenetic phenomena in the regulation of *CTGF*. An inverse correlation between DNA methylation at the *CTGF* locus and *CTGF* gene expression has been demonstrated in ovarian and liver cancers (Kikuchi et al. 2007; Chiba et al. 2005). Furthermore, a recent study investigating the methylome in a large panel (n=367) of haematological neoplasms, identified hypomethylation at the *CTGF* locus, in 42 pre-B ALL specimens and hypermethylation of this locus in 18 T-ALLs, as well as 54 diffuse large B-cell lymphomas (DLBCL) (Martin-Subero et al. 2009). This is compelling evidence for the involvement of epigenetic phenomena in deregulation of the *CTGF* locus, however the high throughput array used in this study only targeted a small number of CpGs (1505 individual CpG sites in 807 genes), and only 2 CpGs were interrogated at the *CTGF* locus. The first CpG designated as P693, was located within the AP-1 binding site of the *CTGF* promoter at position -638, and was de-methylated in pre-B ALL specimens. The second site designated as E156 was located within exon 1 at position +212 and was methylated in T-ALL and DLBCL specimens (Martin-Subero et al. 2009). The authors of this study did not investigate whether *CTGF* gene expression correlated with the differential methylation of these CpG residues. Collectively, these studies point to a significant role for epigenetic regulation of *CTGF* expression. Thus, further more detailed investigation into the epigenetic state of the *CTGF* locus in leukaemia is warranted.

A comprehensive assessment of the methylation status of the *CTGF* locus was performed in cell lines and primary specimens to examine if aberrant CpG methylation may contribute to deregulation of the *CTGF* locus in pre-B ALL. The specific aims of the experiments presented in this chapter were as follows: 1) to investigate whether the

methylation status of the *CTGF* locus correlated with *CTGF* mRNA levels in cell lines, 2) to investigate if such a correlation existed in primary patient specimens, 3) and finally to examine whether modulating the global epigenetic state of cells *in vitro* could influence *CTGF* gene expression.

6.2 Results

6.2.1 Identification of a CpG island at the human *CTGF* genomic locus

To evaluate the distribution of CpG dinucleotides at the *CTGF* locus, the promoter and coding region from -800 to +3200 (relative to the TSS) was scrutinized using the “CpG island Searcher” program <http://www.uscnorris.com/cpgislands2/cpg.aspx> (Takai & Jones 2003). This program can detect an overrepresentation of CpG dinucleotides (CpGs) in DNA sequences defined by the user. The analysis identified a CpG island of approximately 1400 bp spanning the *CTGF* proximal promoter and extending to within exon 3 (Figure 6.1). Closer inspection revealed that the region containing the highest density of CpGs spanned position -91 to +883, comprising 111 individual CpG residues beginning around 100bp upstream from the canonical TSS and ending at the 5' end of exon 3. The identified *CTGF* CpG island is shown in more detail in Figure 6.2.

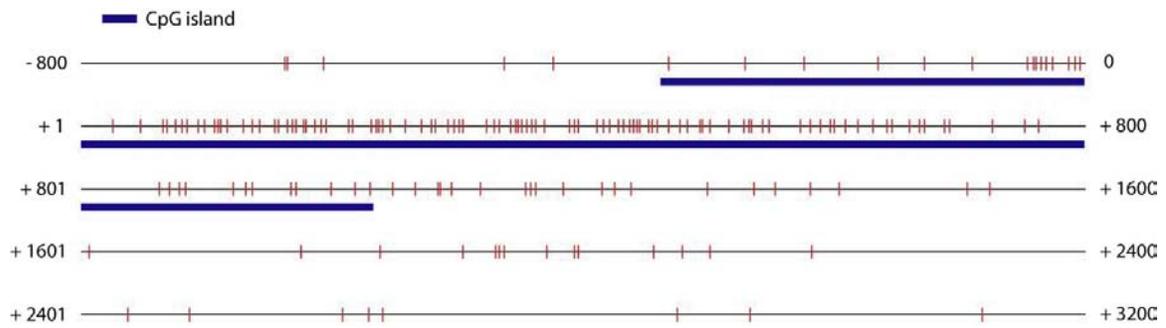


Figure 6.1 CpG dinucleotide distribution within the *CTGF* locus

CTGF genomic DNA sequence from -800 to +3200 was examined for distribution of CpG dinucleotides using the online tool “CpG Island Searcher”(Takai & Jones 2003). Nucleotide positions are indicated relative to the transcription start site (TSS). Red vertical lines denote identified CpGs. The blue bar indicates the extent of a CpG island within the *CTGF* locus.

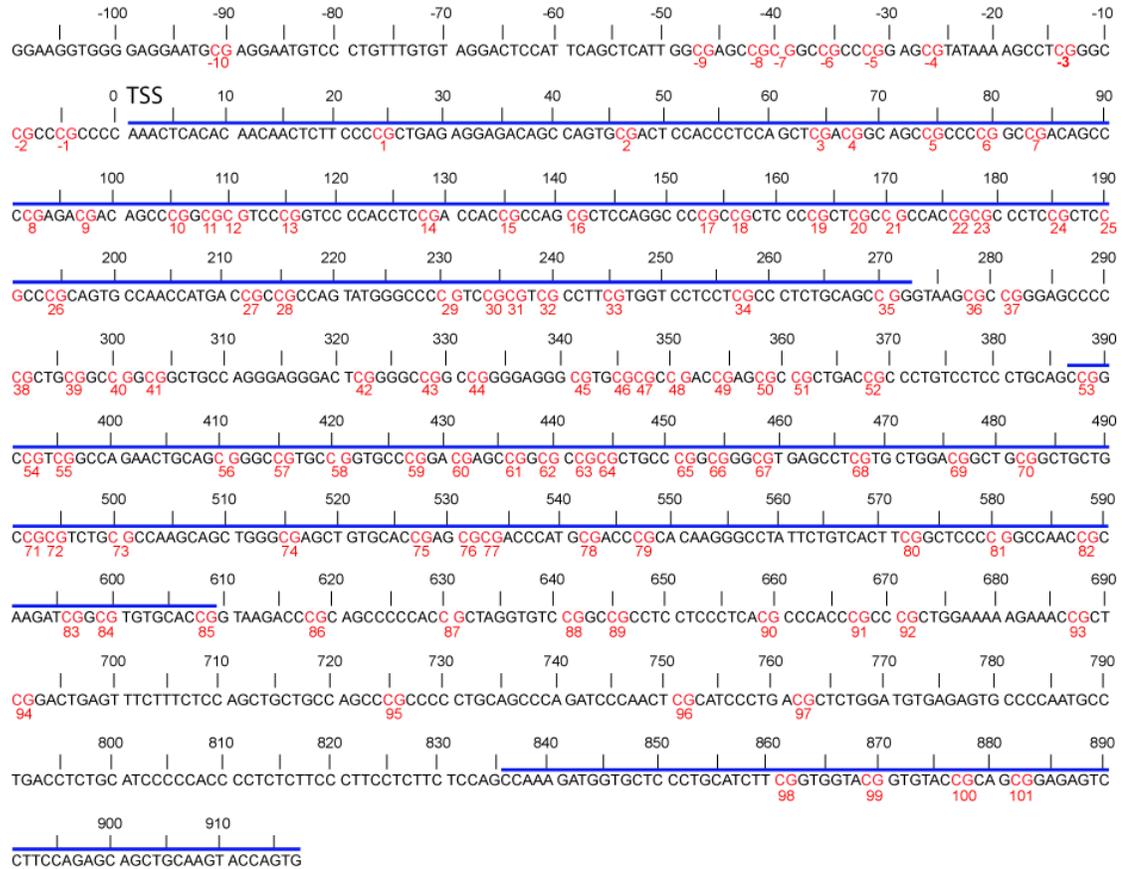


Figure 6.2 Nucleotide sequence of the *CTGF* CpG island

The *CTGF* CpG island extends from position -91 to +883 of the *CTGF* genomic locus. Blue bars denote exons 1 to 3. CpGs are shown in red and are numbered based upon position relative to the TSS.

6.2.2 Methylation correlates with *CTGF* expression in ALL cell lines

To determine if there was a correlation between CpG methylation and *CTGF* gene expression, methylation-specific PCR (msPCR) was performed on bisulfite-treated DNA from eight ALL cell lines, whose *CTGF* status had been established previously by qRT-PCR (see Chapter 4). This panel of cell lines included seven B-lineage ALL cell lines, four of which were positive for *CTGF* mRNA, as well as the T-ALL cell line (JURKAT) that was used as a negative control due to the absence of *CTGF* expression in T-ALLs (Boag et al. 2007). The primers used in the msPCR assays were originally described by

Chiba and colleagues to investigate the methylation status of the *CTGF* locus in hepatoma specimens (Chiba et al. 2005). Figure 6.3A outlines the design of the msPCR reactions. DNA was treated with bisulfite which converts cytosines to uracil via a deamination reaction. Methyl-cytosines were protected from conversion and site-specific primers allowed discrimination between methylated (protected) or unmethylated (converted) DNA using the methylated-site PCR (MSP) and unmethylated-site PCR (USP) primer pairs. These two primer pairs annealed to and allowed amplification from CpGs 5 to 8 at the forward priming site and 27 to 29 at the reverse priming site as outlined in Figure 6.3A. based upon their methylation status, generating a 159 bp PCR product.

The expression level of *CTGF* mRNA in the ALL cell lines is shown in Figure 6.3B. All four of the *CTGF* positive cell lines PER-145, PER-278, PER-371 and PER-377, were unmethylated according to the msPCR, as evidenced by a positive amplification with the USP primer pair only, shown in Figure 6.3C. In contrast, the cell lines that did not express *CTGF*, PER-485, PER-490, PER-495 and JURKAT, were amplified by the MSP primer pair indicating they were methylated at the MSP priming sites.

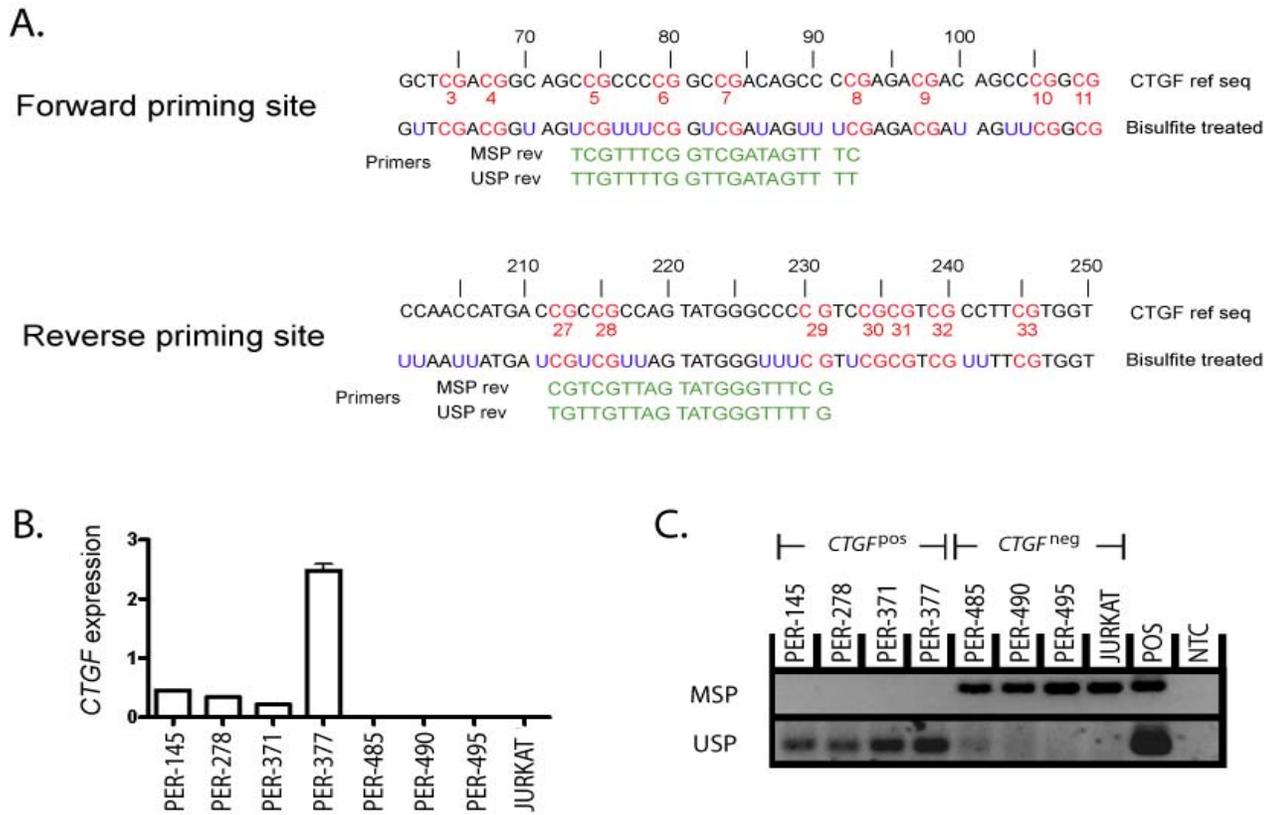


Figure 6.3 Methylation-specific PCR analysis of ALL cell lines

A. *CTGF* reference sequence (ref seq) is shown with nucleotide position relative to the canonical TSS. Unmethylated cytosines are converted to uracil by bisulfite, shown as blue residues in bisulfite-treated sequence. The msPCR priming sites cover CpGs 5 to 8 (numbered in red) at the forward priming site and CpGs 27 to 29 at the reverse priming site. Specific priming by MSP or USP PCR primers (green) infers the methylation status of DNA prior to bisulfite conversion. **B.** *CTGF* expression in a panel of seven B-lineage ALL and one T-ALL cell line, measured by qRT-PCR and normalized to *ACTB*. Error bars indicate the standard error of the mean of technical replicates (n=3). **C.** Cell lines examined by msPCR grouped by *CTGF* mRNA status (positive or negative). For each of the two assays the respective bisulfite-treated positive control was included, as well as a non-template control (NTC). For the USP primers, the BAC clone RP11-69I8 which contains the *CTGF* locus was used while for the MSP primers, the RP11-69I8 BAC was methylated using the SSI methyltransferase prior to bisulfite-treatment.

6.2.3 Bisulfite sequencing of B-lineage ALL cell lines

While the msPCR experiments suggested an association between CpG methylation and reduced *CTGF* expression, a drawback of msPCR is that it restricts examination of the DNA methylation status to only those CpGs targeted by the msPCR primers. Therefore, to gain an understanding of DNA methylation across the entire *CTGF* CpG island, bisulfite sequencing was performed on five of the B-lineage ALL cell lines, three *CTGF*^{pos} (PER-145, PER-278, PER-377) and two *CTGF*^{neg} (PER-485, PER-495). The CpG island was examined as three overlapping amplicons as reported previously (Kikuchi et al. 2007). Amplicon 1 contained nucleotides -113 to +44, Amplicon 2 contained +25 to +566 and Amplicon 3 contained +547 to +916 (numbering relative to TSS). These PCR amplicons were generated from bisulfite-treated genomic DNA, then gel extracted, cloned and sequenced. For each amplicon, eight clones were sequenced in order to gain a representative snapshot of CpG methylation. The results of bisulfite sequencing are shown in Figure 6.4. Consistent with the results of the msPCR, the *CTGF* CpG island was found to be hypomethylated in those cell lines that aberrantly expressed *CTGF* mRNA. The cell line with the highest *CTGF* expression, PER-377, was virtually devoid of CpG methylation at the *CTGF* locus. There was some residual methylation at the 3' end of the CpG island in the lower *CTGF*-expressing cell lines PER-145 and PER-278, with PER-278 showing more pronounced methylation at CpG number 98. In contrast to these findings, the two *CTGF*-negative cell lines PER-485 and PER-495, displayed a high degree of CpG methylation across the *CTGF* locus. This was more pronounced in PER-495, which has only one copy of *CTGF* (see Chapter 4), where the locus was completely methylated.

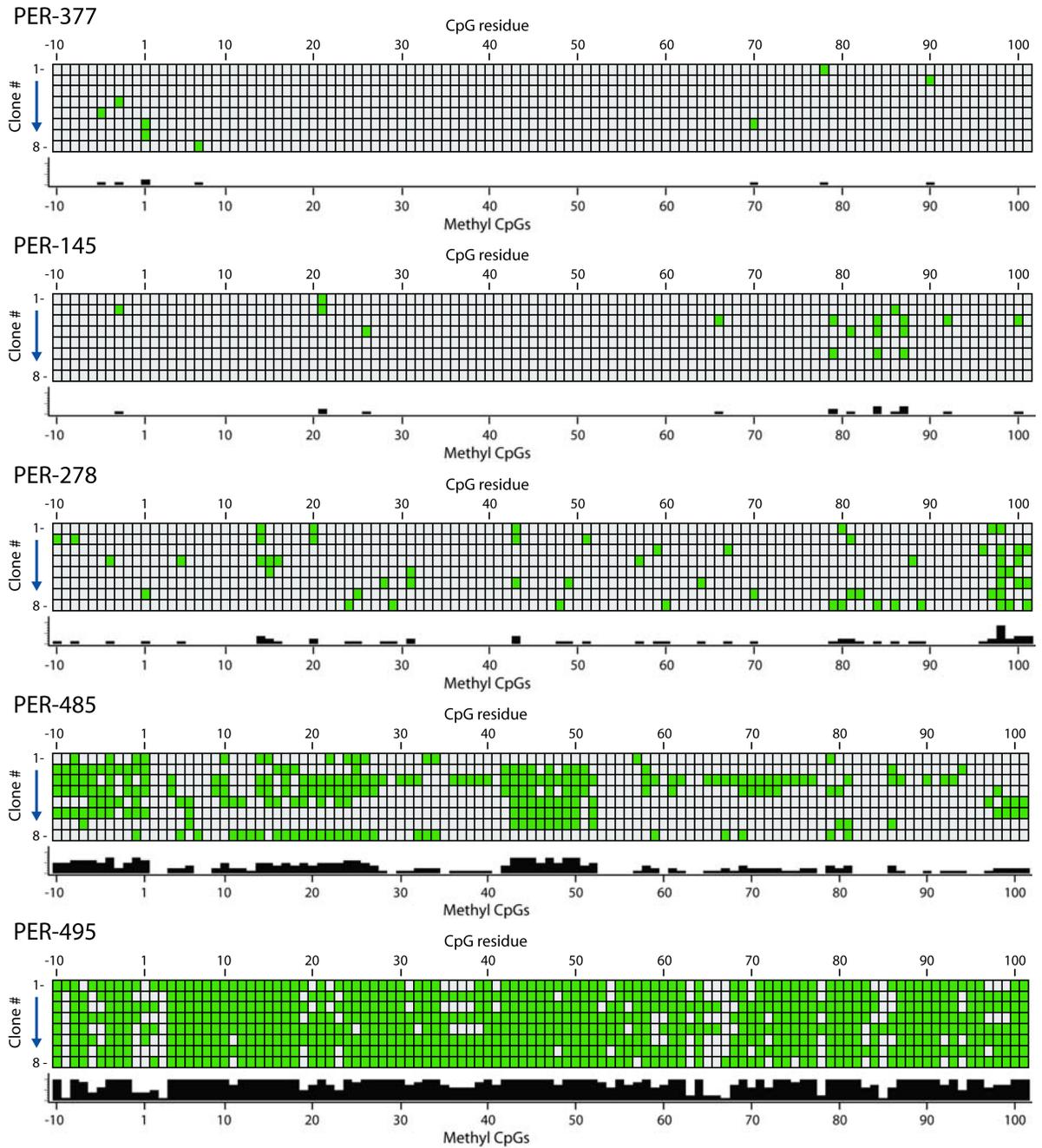


Figure 6.4 Methylation status determined by bisulfite sequencing of the *CTGF* locus correlates with *CTGF* expression in B-lineage ALL cell lines (Figure legend over page)

Figure 6.4 Methylation status determined by bisulfite sequencing of the *CTGF* locus correlates with *CTGF* expression in B-lineage ALL cell lines

The five cell lines examined by bisulfite sequencing were PER-377 (*CTGF*^{high}), PER-145 (*CTGF*^{low}), PER-278 (*CTGF*^{low}), PER-485 (*CTGF*^{neg}) and PER-495 (*CTGF*^{neg}). For each diagram, cell line name is indicated in the top left. Clone numbers 1 to 8 run from top to bottom for each amplicon. The three bisulfite sequencing PCR amplicons are shown as one block from left to right encompassing CpGs -10 to +1, +2 to +78 and +79 to +101 respectively. Methylated CpGs are represented by shaded boxes. The incidence of methylation at each CpG residue is represented in the histogram at the base of each diagram.

6.2.4 Hypomethylation of the *CTGF* locus in primary pre-B ALL

Two independent techniques (msPCR and bisulfite sequencing) demonstrated a correlation between the degree of CpG methylation and expression of *CTGF* in a panel of B-lineage ALL cell lines. To determine if this relationship also existed in primary patient specimens, a panel of pre-B leukaemic bone marrow samples was selected for methylation analysis. Eight samples were analysed using the same msPCR approach that was used to screen the B-lineage ALL cell lines. Of these eight primary samples, qRT-PCR analysis of *CTGF* expression revealed that five exhibited low to negative *CTGF* mRNA expression (*CTGF*^{low/neg}), while three were positive for *CTGF* mRNA at varying levels, collectively referred to herein as *CTGF*^{high} (Figure 6.5A). Despite the heterogeneous levels of *CTGF* expression, all of the primary specimens were amplified by the USP assay only, and thus were classed as unmethylated (Figure 6.5B).

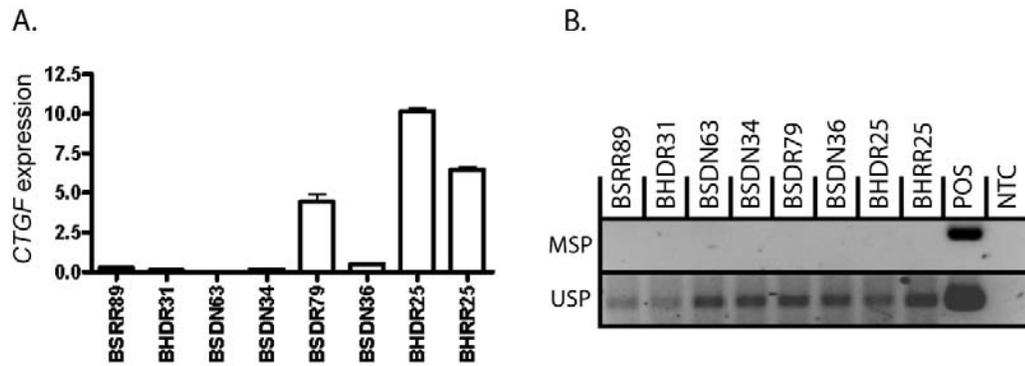


Figure 6.5 msPCR analysis of primary pre-B ALL specimens

A. *CTGF* expression in a panel of primary paediatric pre-B-ALL specimens measured by qRT-PCR and normalized to *ACTB*. Error bars indicate the standard error of the mean of technical replicates (n=3). **B.** Analysis of the methylation status of primary specimens by msPCR. For each of the two assays the respective bisulfite-treated positive control was included, as well as a non-template control (NTC). For the USP primers, the BAC clone RP11-69I8 which contains the *CTGF* locus was used while for the MSP primers, the RP11-69I8 BAC was methylated using the SSI methyltransferase prior to bisulfite-treatment.

6.2.5 Bisulfite sequencing of primary pre-B ALL specimens

The msPCR assay indicated that all eight primary B-ALL specimens tested were unmethylated at the msPCR priming sites, regardless of *CTGF* expression status. However, the methylation status of the rest of the CpG island was unclear. Bisulfite sequencing was therefore undertaken using a larger panel of 14 primary specimens exhibiting varying levels of *CTGF* expression according to microarray data obtained previously (Boag et al. 2007). *CTGF* mRNA levels were validated using qRT-PCR (shown in Figure 6.6). Six of the samples displayed high levels of *CTGF* mRNA (*CTGF*^{high}). The remaining eight samples were classed as *CTGF*^{low/neg}, as seven samples displayed extremely low levels of *CTGF* mRNA and one sample, BSDN63 was completely negative for *CTGF* mRNA after 40 cycles of amplification. Bisulfite sequencing data for the *CTGF*^{high} specimens is shown in below in Figure 6.7 while the *CTGF*^{low/neg} samples are shown in Figure 6.8.

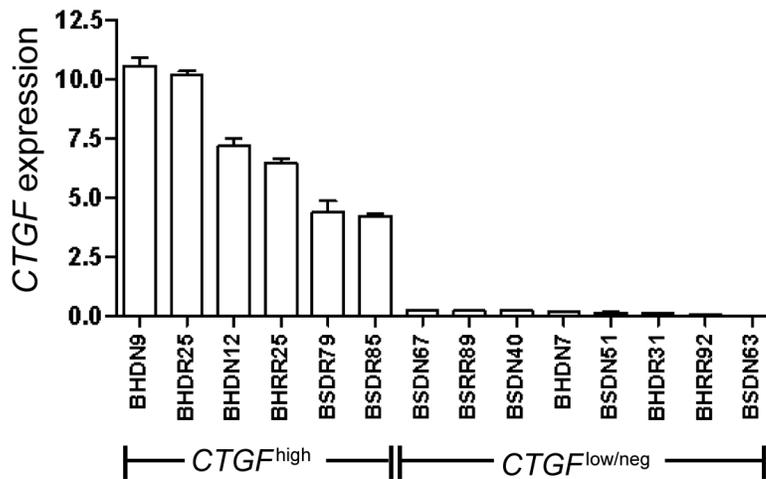


Figure 6.6 *CTGF* mRNA expression in primary pre-B ALL specimens used for bisulfite sequencing

CTGF expression was measured in primary pre-B ALL specimens by qRT-PCR normalized to *ACTB*. Error bars represent the standard error of the mean of technical replicates (n=3). Samples were classed as either $CTGF^{high}$ or $CTGF^{low/neg}$.

Bisulfite sequencing revealed that all of the $CTGF^{high}$ specimens were hypomethylated across the *CTGF* CpG island. However, in contrast to the inverse correlation observed in cell lines between CpG methylation and *CTGF* expression, this correlation was not found in primary specimens, as none of the $CTGF^{low/neg}$ primary specimens displayed significant levels of DNA methylation across the *CTGF* locus. One exception was BHDN7 which displayed methylation of CpGs 15 to 19, 21 to 26, 31 to 33 and 36 to 44 in two clones, suggesting mono-allelic methylation of these CpG residues. There was some sporadic methylated CpGs in BHDR31 which was methylated at the 3' end of Amplicon 3 and BHRR92 was methylated at CpG 24, however no single CpG residue was consistently methylated amongst replicates, indicating that CpG methylation was unlikely to contribute to silencing of *CTGF* expression in these specimens. These results indicate that hypomethylation of the *CTGF* locus is a common feature of primary pre-B ALL specimens, irrespective of *CTGF* expression status.

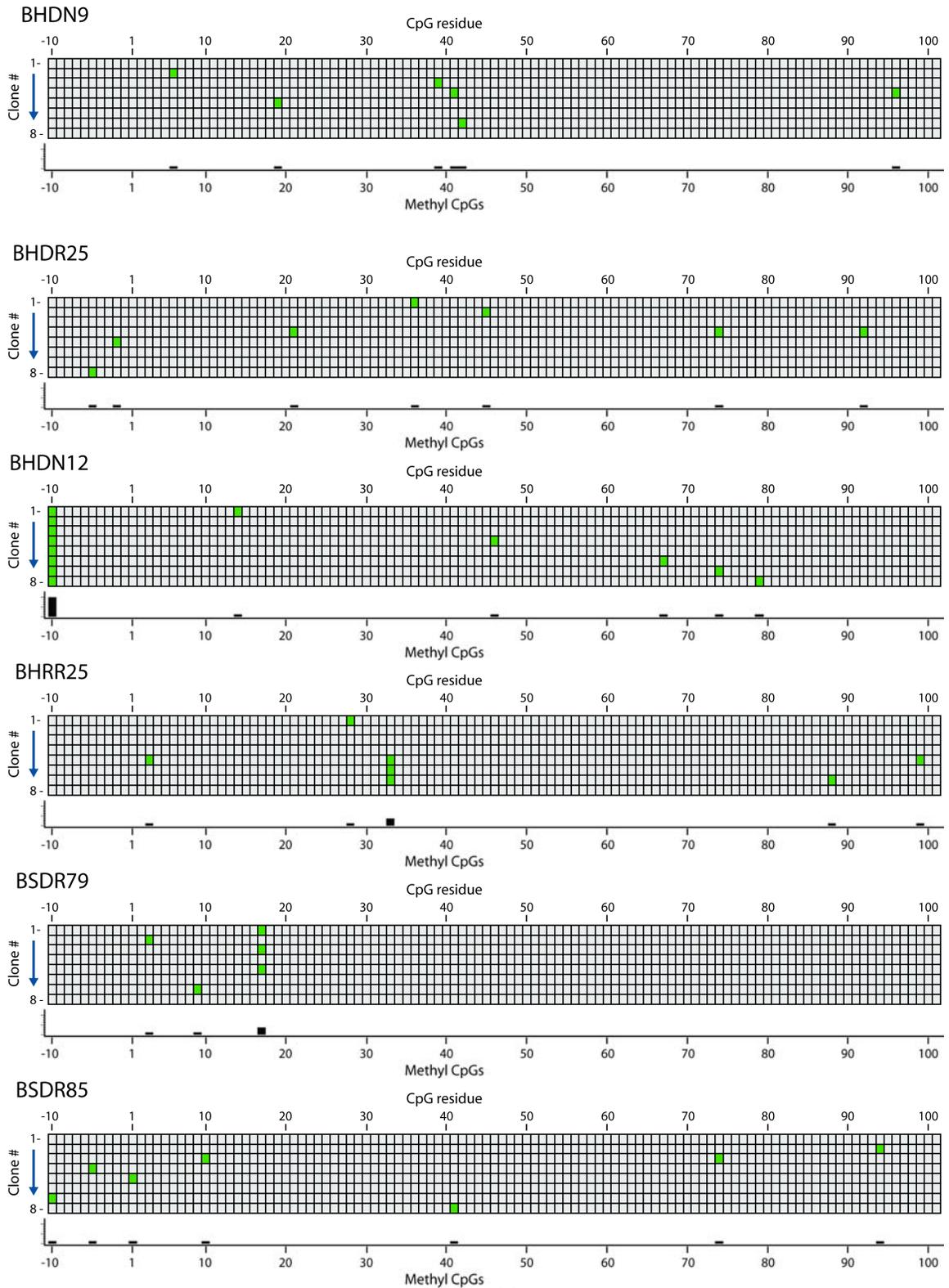


Figure 6.7 Bisulfite sequencing of *CTGF*^{high} primary pre-B ALL specimens (Figure legend over page).

Figure 6.7 Bisulfite sequencing of *CTGF*^{high} primary pre-B ALL specimens
(Continued from previous page)

For each diagram, specimen name is indicated in the top left. Clone numbers 1 to 8 run from top to bottom for each amplicon. The three bisulfite sequencing PCR amplicons are shown as one block from left to right encompassing CpGs -10 to +1, +2 to +78 and +79 to +101 respectively. Methylated CpGs are represented by shaded boxes. The incidence of methylation at each CpG residue is represented in the histogram at the base of each diagram

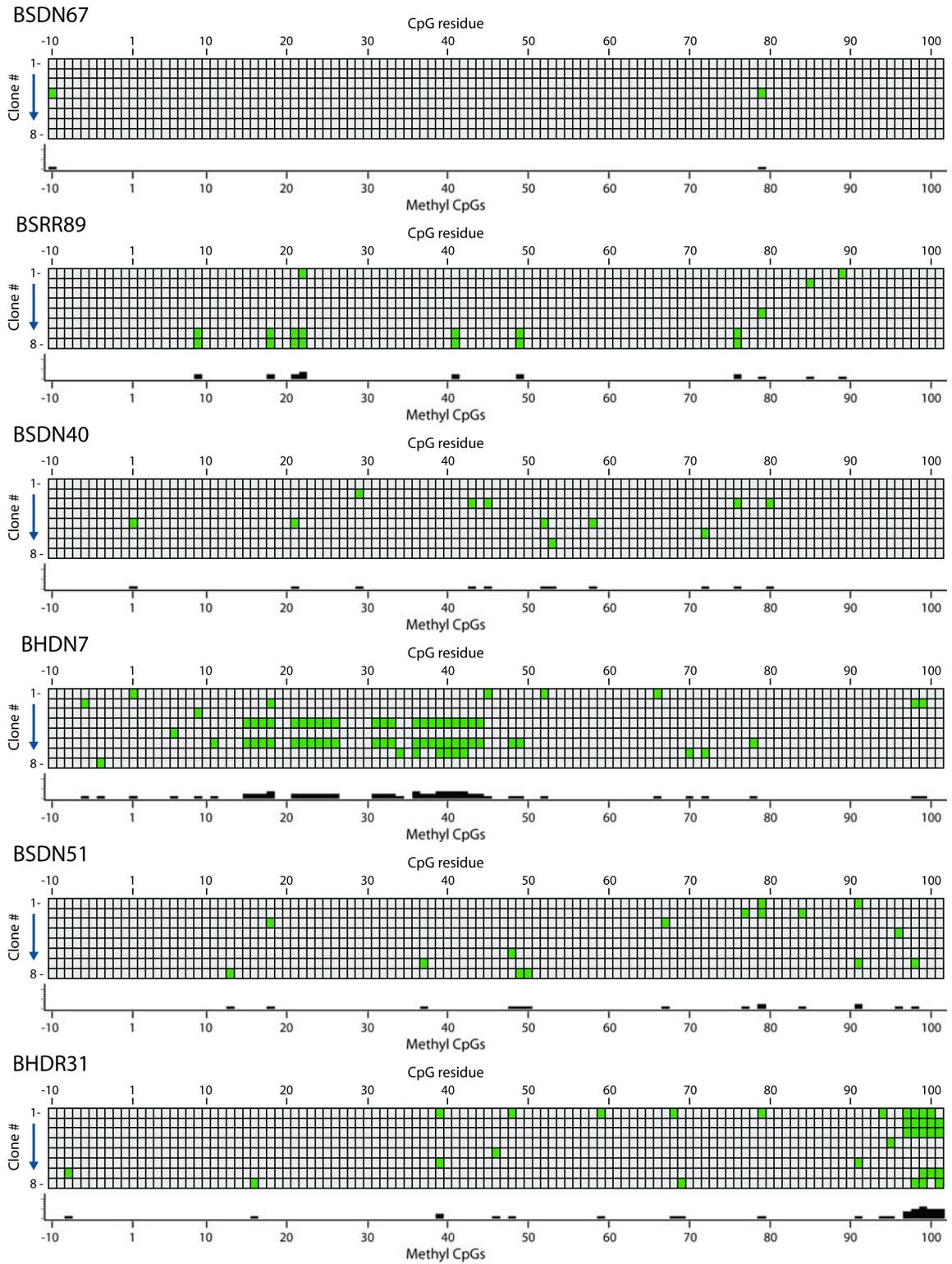


Figure 6.8 Bisulfite sequencing of *CTGF*^{low/neg} primary pre-B ALL specimens

(Figure and legend continued over page).

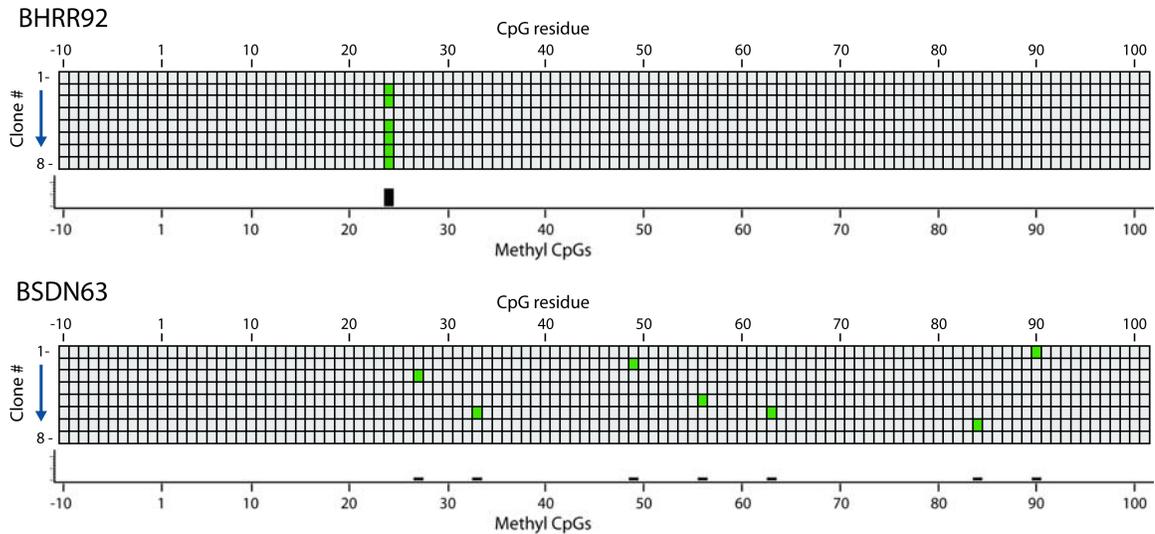


Figure 6.8 Bisulfite sequencing of *CTGF*^{low/neg} primary pre-B ALL specimens
(Continued from previous page)

For each diagram, specimen name is indicated in the top left. Clone numbers 1 to 8 run from top to bottom for each amplicon. The three bisulfite sequencing PCR amplicons are shown as one block from left to right encompassing CpGs -10 to +1, +2 to +78 and +79 to +101 respectively. Methylated CpGs are represented by shaded boxes. The incidence of methylation at each CpG residue is represented in the histogram at the base of each diagram

6.2.6 Bisulfite sequencing of primary T-ALL and CD34^{pos} specimens

Microarray experiments had previously revealed that *CTGF* was not expressed in T-ALL specimens, nor CD34^{pos} cells enriched from normal bone marrow (Boag et al. 2007; Boag et al. 2006). Analysis of T-ALL specimens by qRT-PCR did not detect *CTGF* mRNA (data not shown), and RNA was not available from CD34^{pos} cells. To examine the methylation status of the *CTGF* locus in these two cell types, DNA from two T-ALL primary bone marrow specimens (THDN2 and THDN6) and the CD34^{pos} cells isolated by Boag and colleagues (Boag et al. 2007), were analysed by bisulfite sequencing of the *CTGF* locus. Figure 6.9 shows the results of bisulfite sequencing of these primary specimens. Both of the T-ALL specimens displayed extensive methylation in Amplicons 1 and 3, while the middle amplicon was not consistently methylated, but

did show focal mono-allelic methylation in both cases. Bisulfite sequencing of THDN2, showed that Amplicon 2 was sporadically methylated at different regions on each allele. In THDN6, Amplicon 2 was hypermethylated on one allele, while the other exhibited methylation of a single CpG dinucleotide (CpG 33). Similar to the primary pre-B ALL specimens, CD34-positive bone marrow cells were hypomethylated across all three bisulfite sequencing amplicons.

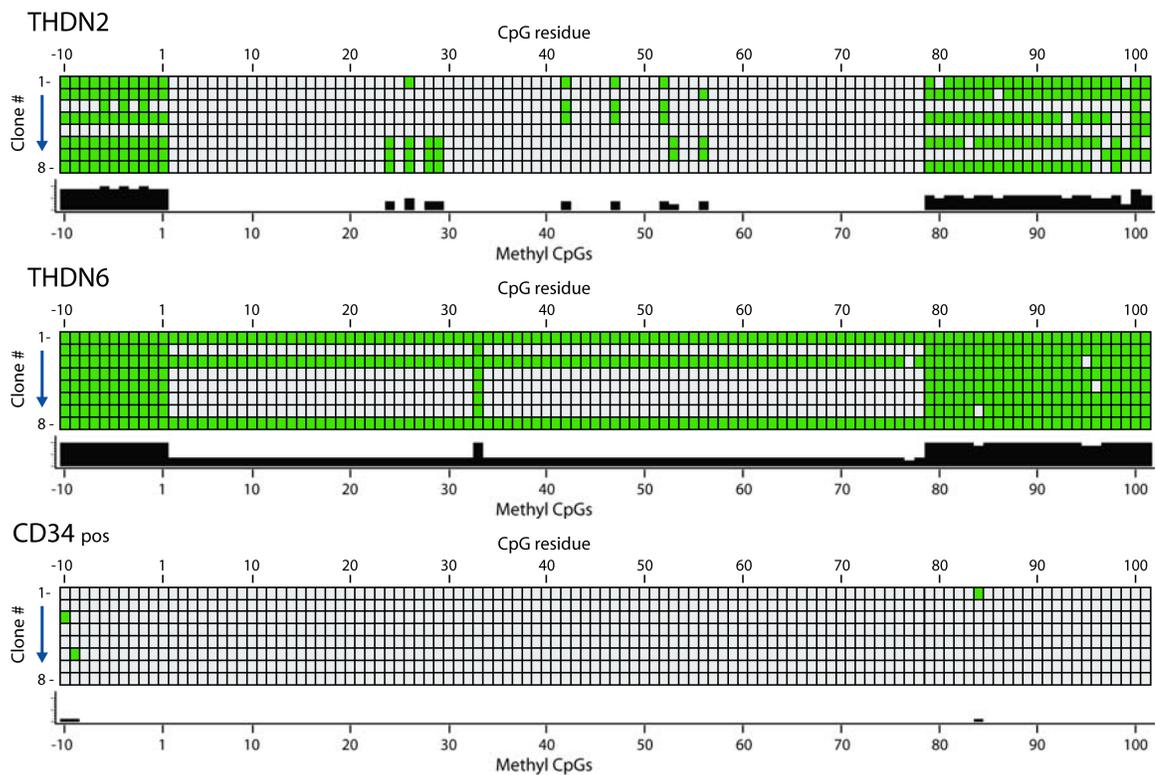


Figure 6.9 Bisulfite sequencing of primary T-ALL specimens and normal CD34 positive bone marrow cells

For each diagram, specimen name is indicated in the top left. Clone numbers 1 to 8 run from top to bottom for each amplicon. The three bisulfite sequencing PCR amplicons are shown as one block from left to right encompassing CpGs -10 to +1, +2 to +78 and +79 to +101 respectively. Methylated CpGs are represented by shaded boxes. The incidence of methylation at each CpG residue is represented in the histogram at the base of each diagram.

6.2.7 Modulation of the global epigenetic state of B-lineage ALL cell lines

Expression of *CTGF* in B-lineage ALL cell lines was highly correlated with the epigenetic status of the *CTGF* locus. Thus, altering the global epigenetic state of ALL cell lines *in vitro* might be expected to lead to changes in *CTGF* gene expression. To test this hypothesis, five B-lineage ALL cell lines were treated with the DNA demethylating agent 5-aza-2'-deoxycytidine (Aza) and the histone deacetylase (HDAC) inhibitor Trichostatin-A (TSA). The three *CTGF*^{neg} cell lines used for this experiment were PER-485, PER-490 and PER-495. In all three cases, treatment with Aza or TSA was highly cytotoxic compared with medium controls and this persisted for two biological replicates (data not shown). Two of these cell lines, PER-490 and PER-485 were derived from infant ALL specimens exhibiting t(4;11) MLL rearrangements which are known to exhibit a marked sensitivity to HDAC inhibitors and Aza (Tonelli et al. 2006; Niitsu et al. 2001). For this reason, these cell lines were not examined further. Therefore two *CTGF*^{pos} cell lines were examined for their response to Aza and TSA, namely PER-145 and PER-377. These two cell lines were shown to be hypomethylated at the *CTGF* locus by bisulfite sequencing, see Figure 6.4.

In the B-ALL cell line PER-377, Aza alone resulted in a modest but not statistically significant increase in *CTGF* expression (Figure 6.10). Treatment with TSA alone also had little effect on *CTGF* expression. However, co-treatment with both agents resulted in a marked and significant increase in *CTGF* expression compared to untreated cells. In the pre-B ALL cell line PER-145, treatment with either agent alone and in combination resulted in a significant increase in *CTGF* expression compared to medium only controls confirming that both DNA methylation and histone acetylation can influence *CTGF* expression in this cell line.

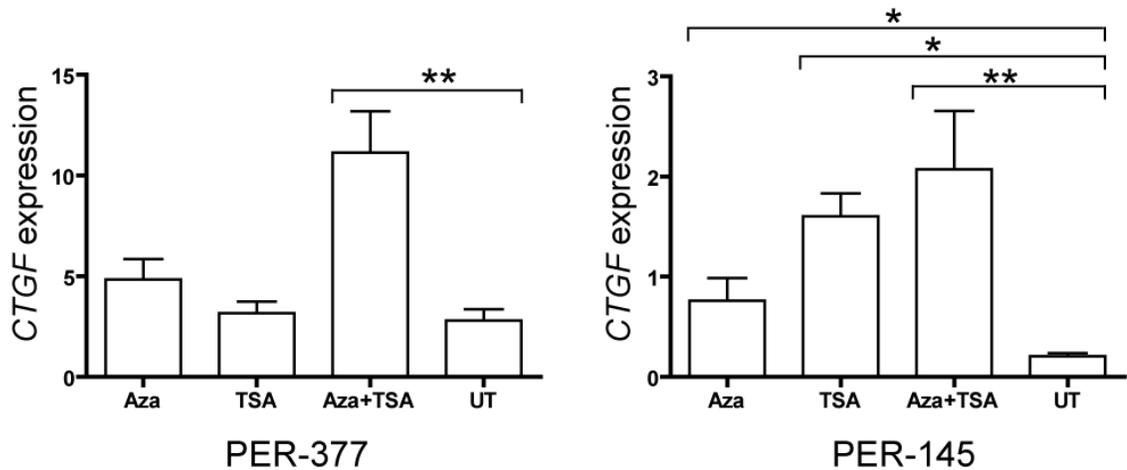


Figure 6.10 The effect of Aza and TSA on *CTGF* expression in B-lineage ALL cell lines PER-377 and PER-145

Cells were treated with 5 μ M 5'aza-2'-deoxycytidine (Aza), 330nM Trichostatin-A (TSA), a combination of both Aza and TSA or culture medium only (UT). Changes in *CTGF* gene expression were measured by qRT-PCR normalized to *ACTB*. Error bars represent the standard error of the mean of three independent experiments. * $p < 0.05$, ** $p < 0.005$.

6.3 Discussion

Investigation of the human genomic *CTGF* locus revealed the presence of a CpG island of approximately 1 kb that was predominantly localised to the 5' end of the *CTGF* gene and coding region. This study showed for the first time that methylation across this region was inversely correlated with *CTGF* gene expression in B-lineage ALL cell lines, thus indicating that this locus can be regulated epigenetically and that hypomethylation, rather than mutation, provides a plausible mechanism for its deregulated expression in pre-B ALL. This finding extended other studies that made use of cell lines derived from hepatocellular carcinoma and ovarian cancers (Chiba et al. 2005; Kikuchi et al. 2007). Several studies have demonstrated that the methylation patterns observed in cancer cell lines closely resemble those observed in the primary tumours from which they were derived (Ueki et al. 2002; Markl et al. 2001; Lind et al. 2004). Furthermore, studies from this laboratory have shown that cell line models can serve as excellent surrogates for

primary tumours (Beesley et al. 2006; Kees et al. 2003). However, novel findings should, where possible, be validated in primary tumour specimens.

6.3.1 The *CTGF* locus is hypomethylated in primary pre-B ALL

Bisulfite sequencing of primary patient specimens provided a detailed map of CpG methylation of the *CTGF* locus in paediatric pre-B ALL specimens and confirmed that hypomethylation of the *CTGF* coding region is a general feature of this disease, regardless of the level of *CTGF* expression, which directly contrasted with observations made in the B-lineage ALL cell lines. When selecting primary bone marrow specimens for bisulfite sequencing analysis, specimens exhibiting a range of *CTGF* expression levels were selected using microarray data obtained in previous studies (Boag et al. 2006; Boag et al. 2007). When these primary specimens were re-examined for *CTGF* expression using qRT-PCR, 6 out of 14 pre-B ALL specimens displayed high levels of *CTGF* mRNA, while seven of the remaining eight *CTGF*^{low/neg} specimens exhibited extremely low *CTGF* mRNA levels that were unlikely to result in synthesis of biologically relevant levels of CTGF protein. In contrast, when the two T-ALL primary specimens and *CTGF*^{neg} B-lineage ALL cell lines, PER-485, PER-490 and PER-495 were examined by qRT-PCR, *CTGF* mRNA was completely undetectable after 40 cycles of amplification, and this corresponded to hypermethylation at the *CTGF* locus. RNA was not available from the CD34^{pos} specimen for this study, however they were considered negative for *CTGF* mRNA based on microarray data (Boag et al. 2007). These findings suggest a model where hypomethylation of the *CTGF* locus leaves this region susceptible to promiscuous transcriptional activation in primary pre-B ALL, while hypermethylation of the *CTGF* locus in T-ALL and *CTGF*^{neg} cell lines renders this region transcriptionally silent.

This observation is supported by previous findings that identified decreased methylation at the *CTGF* locus in pre-B ALL specimens compared to T-ALL and DLBCL (Martin-Subero et al. 2009). The assertion that the *CTGF* locus is, by contrast, hypermethylated in T-ALL is also supported by Martin-Subero and colleagues, although analysis of a

greater number of patient specimens is required to confirm this definitively. With the exception of CD34^{pos} cells, normal haematopoietic precursor cells were not available for analysis during this study. However, given the observation that both DLBCL and T-ALL cells were hypermethylated at the *CTGF* locus, and conversely this region was hypomethylation in bone marrow derived CD34^{pos} cells, and pre-B ALL cells, it is conceivable that that epigenetic silencing of the *CTGF* locus may occur during normal haemopoiesis. Careful examination of the methylation status of the *CTGF* locus throughout haemopoiesis should be undertaken to determine if and when epigenetic silencing of the *CTGF* locus occurs.

6.3.2 The importance of histone modifications

The heterogeneity in *CTGF* expression between the *CTGF*^{high} and *CTGF*^{low/neg} primary pre-B ALL specimens may also be a result of histone modifications such as methylation or acetylation which can alter chromatin accessibility and gene expression (Smith & Shilatifard 2011). This hypothesis was supported by the observation that *CTGF* expression was enhanced in response to treatment with the HDAC inhibitor TSA in the PER-145 cell line, which already exhibited moderate *CTGF* expression. Reinforcing this finding, HDAC inhibitors have been shown previously to up-regulate *CTGF* expression in renal epithelial cells, hepatoma cell lines and primary hepatocytes (Komorowsky et al. 2009; Chiba et al. 2004). Care needs to be taken when interpreting these results however, as pharmacological inhibition of HDACs is a blunt instrument in investigating the role of histone acetylation in gene regulation and changes in acetylation cannot be targeted to specific loci. A more detailed analysis of the role of histone acetylation in the regulation of the *CTGF* locus could be undertaken by performing chromatin immunoprecipitation (ChIP) using antibodies against acetylated H3 and H4 histones, and comparing *CTGF*^{high} and *CTGF*^{low/neg} pre-B ALL specimens by ChIP-sequencing or ChIP-on-Chip (Ho et al. 2011). The enormous power of next generation sequencing means that global patterns of histone acetylation in pre-B ALL could be elucidated, and this may prove useful for not just analysis of the *CTGF* locus, but other regions of interest in pre-B ALL.

A commonly held view of the role of CpG islands is one where methylated CpGs can direct repression of gene transcription by recruiting methyl-CpG binding proteins such as the methyl-CpG binding domain (MBD) proteins 1 to 4 to influence gene expression (Feng & Zhang 2001). These proteins prevent gene transcription by inhibiting the binding of transcription factors to DNA regulatory elements or by recruiting histone modifying complexes to alter the local chromatin structure. For example, MBD1 recruits the histone methyltransferase SETDB1 which catalyses tri-methylation of lysine 9 on histone H3 (H3K9me3), which is a repressive mark (Snowden et al. 2002; Sarraf & Stancheva 2004). Recent findings however suggest that CpG islands can also be interpreted through the specific recognition of non-methylated CpGs and furthermore that non-methylated CpG islands can reinforce active chromatin marks within a coding region.

Two important histone modifying enzymes, KDM2A and CFP1, have been shown to specifically bind to non-methylated CpGs through their zinc finger (ZF)-CxxC domain (Blackledge et al. 2010; Lee & Skalnik 2005). KDM2A is a histone demethylase and catalyses the removal of repressive dimethyl marks from H3 lysine 36 (H3K36me2) (Li et al. 2009a; Tsukada et al. 2006). ChIP assays demonstrated that while intergenic regions show constant high levels of H3K36Me2, non-methylated CpG islands occupied by KDM2A were depleted of this repressive mark (Blackledge et al. 2010). CFP1 also contains a ZF-CxxC domain and specifically binds non-methylated CpGs. CFP1 associates with the SET1 complex to catalyse the addition of the trimethyl marks to H3 lysine 4 (H3K4me3) (Lee & Skalnik 2005). H3K4me3 is generally associated with the 5' end of gene coding regions as is strongly associated with transcriptional activation (Okitsu & Hsieh 2007). These data clearly demonstrate that these ZF-CxxC proteins bind to non-methylated DNA and enforce active chromatin marks and encouraging transcription of target genes. Determining the array of histone modifications at the *CTGF* locus may shed light on mechanisms controlling gene expression through alterations in chromatin structure. This is no small task due to the large number of histone modifications that can be examined, however this should be considered a

worthwhile goal for future investigation into the role of epigenetic modification in regulating the *CTGF* locus.

6.3.3 Conclusions

In summary, the experiments described in this chapter demonstrated that the *CTGF* locus contains a CpG island extending from the proximal promoter into exon 3, that is hypomethylated in primary paediatric pre-B ALL and thereby provides a mechanism to explain the tendency of this locus to be deregulated in pre B-ALL. Furthermore, analysis of T-ALL specimens and bone marrow derived CD34^{pos} cells points to the potential developmental methylation (and silencing) of the *CTGF* locus during lymphopoiesis, although more evidence is required before any definitive conclusions can be drawn. Modulation of global methylation and histone acetylation influenced *CTGF* expression in B-lineage ALL cell lines and further investigation into the covalent modification of histones occupying *CTGF* associated nucleosomes is warranted. Such an investigation may reveal differences in chromatin state between pre-B ALL cells exhibiting non-methylated CpG island but different levels of *CTGF* expression.

Chapter 7

General Discussion

Chapter 7

General Discussion

7.1 Introduction

The prognosis for patients diagnosed with ALL is excellent compared to other cancers, achieving cure rates of 85-90% for some forms of this disease, yet ALL continues to be the most frequently diagnosed cancer in children worldwide (Pui et al. 2010; Gaynon et al. 2010; Conter et al. 2010). Moreover, while these cure rates are remarkable, around 20% of children diagnosed with ALL still relapse (Conter et al.; Gaynon et al. 2010; Pui et al. 2010), and improving therapy for these patients is of paramount importance. Advances in long-term survival for ALL patients have come about largely through the careful optimisation of treatment protocols, facilitated by large, multi-centre clinical trials (Gaynon et al. 2010; Pui et al. 2010), and not through the development of novel therapeutic agents. One of the most critical aspects of conventional ALL therapy is stratifying patients based on clinical features present at diagnosis, and directing these patients to the appropriate level of treatment (Pieters & Carroll 2010). The aim is to identify patients at a high risk of relapse, and ensure these patients are directed to intensive chemotherapy, to afford them the best chance of achieving a clinical remission. At the same time, low or standard-risk patients are spared from the more toxic therapies reserved for high-risk patients or those that relapse. These treatments can carry significant long-term health risks, and thus limiting their use in children is desirable (Dowling et al. 2010).

However, a key question remains in terms of what biological features can best distinguish these high-risk patients from others and how therapy can be tailored further for high-risk patients according to specific biological features, such as gene expression signatures or clinical features. Clinical trials alone cannot answer these questions, thus the key to improving patient outcomes further is to elucidate the molecular mechanisms underlying ALL. This may facilitate the discovery of novel prognostic biomarkers with

therapeutic relevance, or the elucidation of tumour specific cellular pathways that can be exploited by new or existing cancer treatments for therapeutic gain.

7.2 Study objectives

This thesis investigated the molecular mechanisms responsible for deregulated expression of the proto-oncogene *CTGF* in childhood pre-B ALL. The literature review highlighted that aberrant *CTGF* expression is a frequent event in pre-B ALL (Vorwerk et al. 2000; Sala-Torra et al. 2007; Boag et al. 2007), as well as numerous solid cancers, however this review also highlighted an incomplete understanding of both the mechanisms regulating the *CTGF* gene locus and its involvement in the pathogenesis of ALL. Hence, this thesis addressed four related research questions. The first hypothesis was that *CTGF* expression may be a result of abnormal transcription factor activation or expression. Therefore, the global gene expression profiles of three, independent, paediatric pre-B ALL cohorts, were compared to identify *CTGF*-correlated patterns of gene expression. The second hypothesis was that *CTGF* expression may exhibit prognostic significance and may be associated with clinical features at the time of the patient's diagnosis. Thus, the association between clinical features, *CTGF* expression and patient survival was investigated in the PMH cohort of paediatric pre-B ALL. The third hypothesis was that aberrant *CTGF* expression may have a genetic basis. To address this, a sequencing and hybridisation approach was employed to investigate the *CTGF* gene locus and *CTGF* mRNA transcripts for lesions that may contribute to its deregulated expression. Lastly, the fourth hypothesis was that *CTGF* expression has an epigenetic basis. Hence, methylation-specific PCRs and bisulfite sequencing was performed to investigate the association between DNA methylation and *CTGF* expression. These experiments were expected to shed light on the mechanisms driving *CTGF* expression in pre-B ALL, as well as highlighting its prognostic relevance.

7.3 Critical findings

7.3.1 A possible role for Ikaros proteins in *CTGF* transcription

To examine whether aberrant transcription factor activation or expression was responsible for transcriptional activation of *CTGF* in pre-B ALL, three independent, pre-B ALL microarray datasets were compared and examined for *CTGF*-correlated genes. Six genes were identified by this approach and by comparing the 5' flanking regions of these six genes to that of *CTGF*, as well as each other, it was expected that common transcription factor binding sites would become evident, implicating individual transcription factors as candidates for functional analysis. Somewhat surprisingly, binding sites for nuclear factors known to regulate the *CTGF* locus in other cell types were not identified in these *CTGF*-correlated genes. Instead, *in silico* analysis uncovered binding sites for the Ikaros protein family members IKAROS and HELIOS, in the 5' flanking sequence of these genes.

The Ikaros family of transcriptional regulators have been associated with both transcriptional activation, as well as silencing of target loci by targeting repressive chromatin changes (Koipally et al. 1999; Rebollo & Schmitt 2003). Both HELIOS and IKAROS have been associated with haematological malignancies and *IKAROS* is frequently deleted or mutated in high-risk pre-B ALL (Kuiper et al. 2010; Mullighan et al. 2009). Thus, the Ikaros family of proteins may be involved in silencing the *CTGF* locus, and a loss of IKAROS or HELIOS function may result in dysregulation of *CTGF* expression. Indeed analysis of a cohort of high-risk pre-B ALL patients that were recently examined for *IKAROS* mutations and deletions (Mullighan et al. 2009), confirmed that *CTGF* expression was significantly associated with *IKAROS* genomic lesions. While these findings suggest that IKAROS and/or HELIOS may play a role in regulating the *CTGF* locus, detailed functional characterisation of the effect of these proteins upon the *CTGF* promoter have not yet been undertaken.

This is the first time IKAROS or HELIOS have been implicated in *CTGF* gene expression. The role of the Ikaros family gene mutations and deletions in dysregulation

of the *CTGF* locus and in pre-B ALL more broadly is warranted. *IKAROS* gene mutations or deletions are highly predictive of relapse (Kuiper et al. 2010; Mullighan et al. 2009), and in-frame deletions within the *IKAROS* coding region can result in the synthesis of inhibitory DN isoforms. However, these may also be generated by alternative splicing (Meleshko et al. 2008), making this a challenging area of research. Compounding this, DN isoforms may interact with other Ikaros family members due to conservation in the dimerisation domain of this protein family. Thus, a broad analysis of the mutation and transcriptional status, as well as the functional capacity of all five Ikaros family members should be performed simultaneously in any future studies and correlated to *CTGF* gene expression.

7.3.2 Clinical importance of *CTGF* expression in pre-B ALL

While *CTGF* expression has been associated with a poor outcome in both adult and high-risk paediatric ALL (Sala-Torra et al. 2007; Kang et al. 2010), it was unclear if specific biological features correlated with *CTGF* mRNA levels. Detailed patient data was available for the PMH cohort, and thus associations between a range of clinically relevant biological features and *CTGF* gene expression was examined. Significantly lower mean *CTGF* expression was observed in patients with enlarged lymph nodes, and there was also a similar trend in patients with enlarged spleens, however this did not reach significance. This observation suggests that *CTGF* expression in leukaemic cells does not promote metastasis to sites outside of the bone marrow. As reported in Chapter 1, *CTGF* expression has been both positively and inversely associated with metastasis (Lin et al. 2005; Li et al. 2007), and thus the tissue context is likely to significantly influence the biological outcome of *CTGF* secretion.

To investigate the associations between *CTGF* expression and patient outcomes, the PMH cohort was stratified based upon *CTGF* expression and examined for RFS and overall survival. There was no association between overall survival and *CTGF* expression, however those patients whose leukaemic blasts were *CTGF*^{POS} had a lower RFS compared to those patients whose blasts were *CTGF*^{low/eg} (71% and 83%

respectively). This association in the PMH cohort did not reach statistical significance, and this may be due to the cohort composition which contained 68% standard-risk and 32% high-risk patients. Nonetheless, these findings confirm reports by others that *CTGF* expression is associated with a poor outcome in pre-B ALL. The mechanisms through which *CTGF* exerts its biological effects in ALL remains unclear, however it is suspected that aberrant *CTGF* gene expression in the bone marrow may result in changes in the microenvironment that create a preferred site of sanctuary for leukaemic cells, and this may have important ramifications in regards to the emergence of drug resistance during therapy.

7.3.3 A model for CTGF-mediated leukaemogenesis

The *CTGF* protein is pleiotropic, however *CTGF* has no known role in lymphocyte biology. It is expressed in numerous tumour types, and thus may confer a biological advantage to leukaemic cells through autocrine or paracrine mechanisms, or by altering interactions with the stromal microenvironment. It is now widely accepted that bone marrow stromal cells facilitate adhesion of leukaemic cells within the bone marrow microenvironment, affording them sanctuary from chemotherapy (Weisberg et al. 2008; Meads et al. 2008). *CTGF* can induce terminal differentiation of bone marrow stromal cells (BMSCs) into fibroblasts (Lee et al. 2010) and these BMSC-derived fibroblasts may exhibit some of the growth-promoting effects exhibited by cancer-associated fibroblasts in other cancer types (Allen & Louise Jones 2011).

This proposed model, where secreted *CTGF* has a paracrine effect upon the bone marrow microenvironment is outlined in Figure 7.1, and is supported by two key unpublished findings generated in the laboratory of Prof Ursula Kees. Firstly, both *CTGF*^{pos} and *CTGF*^{neg} pre-B ALL cells fail to respond to treatment with exogenous recombinant human *CTGF* protein *in vitro*, seemingly ruling out autocrine growth effects upon these leukaemic cells. Secondly, pre-treatment of human bone marrow stromal cells (HS5 cell line) with rh*CTGF* resulted in enhanced adhesion, migration, proliferation and most notably drug-resistance, of pre-B ALL cells in subsequent co-

culture experiments. It could be speculated that a paracrine growth loop exists that is facilitated by CTGF-induced changes in stromal cell gene expression. These unpublished findings, which were not part of the present study, may account for the enhanced bone marrow fibrosis observed in pre-B ALL (Noren-Nystrom et al. 2008). This model also provides a plausible mechanism by which CTGF expression in the bone marrow may promote drug resistance and relapse in patients, thus highlighting the importance of better understanding the mechanisms driving its deregulated expression. These studies are ongoing and are expected to yield important data regarding the paracrine effects of CTGF in the bone marrow microenvironment.

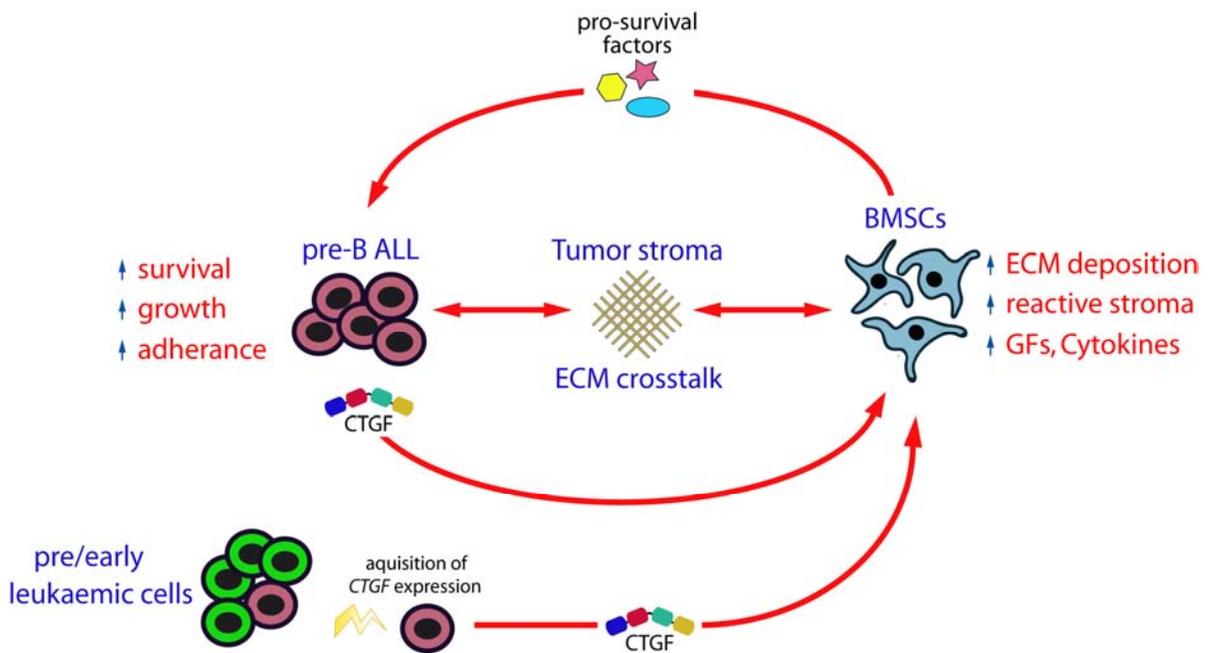


Figure 7.1 Proposed model of the growth-promoting effects of CTGF in the bone marrow microenvironment

Acquisition of *CTGF* gene expression by pre-leukaemic or early leukaemic cells results in aberrant secretion of CTGF protein. Bone marrow stromal cells (BMSCs) respond to CTGF protein by releasing pro-survival signals such as growth factors and cytokines as well as increasing synthesis of extracellular matrix (ECM) proteins. BMSC derived pro-survival factors promote enhanced proliferation and adherence of pre-B ALL cells. Tumour stroma may also facilitate crosstalk between ALL cells and other stromal cells in the bone marrow microenvironment.

7.3.4 Genomic lesions did not contribute to *CTGF* expression

The third major hypothesis of this study was that aberrant *CTGF* expression was the result of genomic lesions affecting the *CTGF* locus. Southern blotting ruled out rearrangement of a 16 kb region encompassing the *CTGF* locus and flanking regions. Similarly, gene copy number analysis revealed that primary pre-B ALL specimens and ALL cell lines were diploid for *CTGF*. Thus gross structural alterations could not account for abnormal *CTGF* expression. Sequencing of the *CTGF* promoter and 3'UTR in ALL cell lines did not identify any mutations that could account for either abnormal promoter activation, or a failure in post-transcriptional regulation of *CTGF* mRNA. It is possible that cryptic enhancer or promoter elements outside of the region sequenced in this study may be responsible for the activation of the *CTGF* locus, and sequencing of upstream regions may yet uncover mutations relevant to *CTGF* expression. It should be noted that the rs6918698 C>G SNP at position -739 was identified in one of the *CTGF*^{pos} ALL cell lines, however given the fact that it was only identified in one of four *CTGF*^{pos} cell lines, it seems unlikely that this SNP contributes significantly to *CTGF* promoter dysregulation. Nonetheless, investigation of the incidence of this SNP in primary pre-B ALL specimens may reveal association with *CTGF* expression, and thus could be considered in future studies if only to rule out its involvement.

7.3.5 *CTGF* mRNA is subject to alternative splicing in pre-B ALL

While evidence of aberrant *CTGF* expression has been provided from several sources (Sala-Torra et al. 2007; Vorwerk et al. 2000; Boag et al. 2007), these findings have all been generated by detection of *CTGF* mRNA using qRT-PCR and microarrays, targeting the 3' UTR of the *CTGF* transcript. Northern blotting and cDNA sequencing was used to evaluate the complement of *CTGF* mRNA transcripts produced in ALL cell lines. These experiments confirmed that while the majority of *CTGF* mRNA (>90%) was the canonical 2.4 kb mRNA, an array of *CTGF* transcripts were generated by alternative pre-mRNA splicing, as well as alternative transcription start site usage in ALL cells. This is the first evidence of alternative splicing of *CTGF* mRNA in any tissue type and thus is an important finding, as some of these *CTGF* spliceforms may exist in

other cancer types or normal tissues and have the potential to encode truncated CTGF proteins as well as synthesizing transcripts that can avoid post-transcriptional control of CTGF expression. Analysis of *CTGF* mRNA splicing in primary pre-B ALL specimens was not performed in the present study. However, such an investigation is warranted to establish accurate patterns of *CTGF* mRNA synthesis in pre-B ALL that may uncover tumour-specific *CTGF* transcripts with altered coding potential or susceptibility to post-transcriptional control mechanisms.

7.3.6 *CTGF* mRNA may not reflect CTGF protein levels in ALL

A disparity was observed between *CTGF* mRNA and protein levels in ALL cell lines. While four cell lines were *CTGF*^{pos}, only the PER-377 cell line exhibiting the highest level of *CTGF* mRNA synthesized detectable levels of CTGF protein, suggesting that post-transcriptional regulation of *CTGF* mRNA may be occurring in these cell lines. This finding suggests that the levels of *CTGF* mRNA detected in primary specimens may not accurately represent the amount of CTGF protein produced by these cells. Therefore, CTGF protein would make a more attractive target as a biomarker and may hold more prognostic significance than mRNA levels alone. Future studies should investigate the correlation between *CTGF* mRNA and protein levels in primary specimens. Furthermore, evaluation of the association between CTGF protein expression and patient outcomes is justified.

If CTGF protein levels are to be examined for prognostic significance in pre-B ALL, this raises the question as to what biological material is most appropriate. Diagnostic bone marrow aspirates are routinely collected to confirm a diagnosis of ALL. These specimens represent the most relevant biological material for detecting CTGF protein levels, as CTGF is predicted to significantly alter the bone marrow microenvironment. Purification of mononuclear cells from diagnostic bone marrow aspirates is usually performed with a ficoll-hypaque density gradient. This protocol could be easily adapted to ensure that the acellular plasma fraction of this density gradient is retained and cryopreserved for subsequent molecular investigation of secreted growth factors present

in the bone marrow space. This method would fit into current diagnostic sample processing protocols without significant procedural changes, however aspirates would need to be obtained using the anticoagulant ethylenediaminetetraacetic acid (EDTA) only, as the other commonly used anticoagulant heparin would bind to CTGF through its CT domain, potentially confounding protein detection. Collection of this material for retrospective analysis should be encouraged to focus not only on CTGF, but also other biomarkers present in the bone marrow that may come to light in future research.

7.3.7 The *CTGF* locus is hypomethylated in pre-B ALL

The final hypothesis of this study was that aberrant *CTGF* expression may be a result of epigenetic changes in ALL cells. DNA methylation and histone acetylation has been shown to play a role in modulating the transcriptional potential of the *CTGF* locus in liver and ovarian cancers (Chiba et al. 2004; Chiba et al. 2005; Kikuchi et al. 2007) and global hypomethylation of cancer cells is an accepted mechanism by which oncogenes can be activated in cancer cells (Feinberg & Vogelstein 1983b; Tsujiuchi et al. 1999; Hanada et al. 1993). A combination of msPCR and bisulfite sequencing was used to investigate DNA methylation at the *CTGF* locus in ALL cell lines and primary specimens. This was the first such detailed analysis of epigenetic regulation of the *CTGF* locus in ALL.

An inverse correlation between *CTGF* gene expression and DNA methylation was observed in a panel of ALL cell lines. In contrast, all fourteen pre-B ALL primary specimens were hypomethylated at the *CTGF* locus despite displaying heterogeneous levels of *CTGF* expression. It is not clear whether the *CTGF* gene locus is normally subjected to epigenetic silencing during lymphoid specification or maturation, however in contrast to pre-B ALLs the *CTGF* locus was hypermethylated in two T-ALL specimens investigated in this study and this finding is supported by others that have observed a similar pattern in global methylation analysis of a large panel of haematological malignancies (Martin-Subero et al. 2009). Thus, hypomethylation of the locus is a conceivable mechanism by which aberrant *CTGF* expression may arise. The

ability of epigenetic factors, including DNA methylation and histone acetylation to modulate *CTGF* expression were confirmed *in vitro*, and this finding is also in agreement with published data from others relating to epigenetic regulation of *CTGF* in other cancer types (Kikuchi et al. 2007; Chiba et al. 2004). Because *CTGF* mRNA is not detectable at any stage in normal haemopoiesis, it is highly likely that this locus is epigenetically silenced during normal development. Careful analysis of the epigenetic state of the *CTGF* locus at each stage of lymphopoiesis is a worthwhile endeavour to confirm whether epigenetic processes are responsible for maintaining this region in a transcriptionally silent state.

7.4 Future directions

7.4.1 Prognostic relevance of CTGF

CTGF mRNA was associated with lower RFS in the PMH cohort of paediatric ALL and unpublished data implicates *CTGF* protein in modulating the bone marrow microenvironment to support leukaemic cells. The prognostic relevance of secreted *CTGF* protein in the bone marrow should be investigated, as there is abundant data supporting associations between aberrant *CTGF* mRNA expression and unfavourable outcome in both paediatric and adult ALL. As discussed, diagnostic bone marrow specimens represent the most appropriate biological material for this analysis as this material would provide a snapshot of protein levels in the bone marrow space. Thus, bio-banking of this material should be encouraged to facilitate retrospective analysis of the associations between patient outcomes and secreted growth factors, including *CTGF* or other factors that are highlighted as important candidates in the pathogenesis of ALL.

7.4.2 Modeling the role of CTGF in ALL pathogenesis in-vivo

Murine models of B-lineage lymphomas or leukaemias, as well as xenographs of primary human pre-B ALLs represent suitable *in vivo* models with which to study the effect of *CTGF* gene expression upon the biology of leukaemia, as well as the bone

marrow microenvironment. Human pre-B ALL murine xenographs provide the opportunity to examine whether inhibition of CTGF activity can affect the viability, engraftment or drug resistance profiles of pre-B ALL cells *in vivo*. In an orthotopic murine model of pancreatic adenocarcinoma, in which *CTGF* is expressed at high levels, administration of a human monoclonal CTGF antibody attenuated tumour growth and metastasis. This study demonstrated that secreted CTGF can have a significant effect upon tumour biology and the surrounding stroma, and further that blocking the function of CTGF can influence tumour growth (Aikawa et al. 2006). This highlights the potential for a similar role in ALL therapy. Preliminary studies have shown that co-incubation of pre-B ALL cells with CTGF-treated bone marrow stromal cells affords a significant level of protection against chemotherapy. Thus, it is conceivable that inhibiting CTGF function using blocking antibodies during induction chemotherapy could hypothetically reduce the emergence of resistance. There is much to be learnt from murine models before such a hypothesis could be tested, however attenuating CTGF function could conceivably provide a novel therapeutic strategy for the management of paediatric pre-B ALL in the future.

7.4.3 Factors interacting with the *CTGF* gene locus

While nuclear factors known to regulate the *CTGF* locus were not expressed at higher levels in *CTGF*^{POS} cells, these factors may be activated post-transcriptionally, or through aberrant activation of binding partners. As such, analysis of nuclear factors bound at the *CTGF* gene locus should be explored by chromatin immunoprecipitation (ChIP). Candidates for analysis by ChIP include those factors known to activate the *CTGF* promoter in other cell types such as AP-1, Ets-1, MMP3 or SP-1 (Yu et al. 2009; Van Beek et al. 2006; Eguchi et al. 2008; Holmes et al. 2003), as well as members of the Ikaros family of nuclear factors highlighted in the present study.

7.5 Final conclusions

This thesis supports previous observations that *CTGF* expression is associated with a poor outcome in paediatric pre-B ALL and identified an association between *CTGF*

gene expression and reduced lymph node metastasis, suggesting *CTGF* expression may influence homing of leukaemic cells. A disparity between *CTGF* mRNA and protein levels was observed, thus future studies should develop methods for detecting *CTGF* protein levels in the bone marrow space. Analysis of global gene expression data from three independent cohorts implicated the Ikaros family lymphoid-specific nuclear factors in regulation of the *CTGF* gene locus. Functional investigations into the role of these lymphoid-specific nuclear factors in altering *CTGF* expression are now warranted. Genomic lesions such as gene rearrangements, copy number alterations, and gene mutations were ruled out as mechanisms contributing to *CTGF* expression in pre-B ALL, however hypomethylation of the *CTGF* locus was a feature of all primary pre-B ALL specimens examined, providing a plausible mechanism for its deregulated expression. As well as aberrant epigenetic marks, alternative splicing of *CTGF* mRNA was observed and minor transcripts were characterized that displayed altered coding potential and others with the ability to evade post-transcriptional control mechanisms.

In conclusions this thesis has improved our understanding of the molecular mechanisms contributing to *CTGF* gene expression in pre-B ALL, as well as highlighting important aspects of *CTGF* biology that require further investigation. Continued research into the role of *CTGF* gene expression in the development and maintenance of pre-B ALL may uncover biological interactions that are critical for ALL pathogenesis. There is evidence emerging from both *in vitro* and *in vivo* models to suggest that inappropriate *CTGF* secretion by leukaemic cells may have significant effects upon the establishment and progression of this disease. Furthermore these interactions may be therapeutically vulnerable and may represent a valid avenue of improving patient therapy in the future. By continuing to advance our understanding of key molecular pathways underlying ALL, we can move toward improving treatments and ultimately long-term outcomes for children diagnosed with this disease.

Chapter 8

References

References

Declaration

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Abreu, J. G., Ketpura, N. I., Reversade, B. & De Robertis, E. M. 2002, 'Connective-tissue growth factor (CTGF) modulates cell signalling by BMP and TGF-beta', *Nat Cell Biol*, vol. 4, no. 8, pp. 599-604.

Agueli, C., Cammarata, G., Salemi, D., Dagnino, L., Nicoletti, R., La Rosa, M., Messina, F., Marfia, A., Bica, M. G., Coniglio, M. L., Pagano, M., Fabbiano, F. & Santoro, A. 2010, '14q32/miRNA clusters loss of heterozygosity in acute lymphoblastic leukemia is associated with up-regulation of BCL11a', *Am J Hematol*, vol. 85, no. 8, pp. 575-8.

Aikawa, T., Gunn, J., Spong, S. M., Klaus, S. J. & Korc, M. 2006, 'Connective tissue growth factor-specific antibody attenuates tumor growth, metastasis, and angiogenesis in an orthotopic mouse model of pancreatic cancer', *Mol Cancer Ther*, vol. 5, no. 5, pp. 1108-16.

Akiyama, Y. n.d., *TFSEARCH: Searching transcription factor binding sites*.

Albihn, A., Johnsen, J. I. & Henriksson, M. A. 2010, 'MYC in oncogenesis and as a target for cancer therapies', *Adv Cancer Res*, vol. 107, pp. 163-224.

Allen, M. & Louise Jones, J. 2011, 'Jekyll and Hyde: the role of the microenvironment on the progression of cancer', *J Pathol*, vol. 223, no. 2, pp. 162-76.

Amati, B., Frank, S. R., Donjerkovic, D. & Taubert, S. 2001, 'Function of the c-Myc oncoprotein in chromatin remodeling and transcription', *Biochim Biophys Acta*, vol. 1471, no. 3, pp. M135-45.

Annino, L., Vegna, M. L., Camera, A., Specchia, G., Visani, G., Fioritoni, G., Ferrara, F., Peta, A., Ciolli, S., Deplano, W., Fabbiano, F., Sica, S., Di Raimondo, F., Cascavilla, N., Tabilio, A., Leoni, P., Invernizzi, R., Baccarani, M., Rotoli, B., Amadori, S. & Mandelli, F. 2002, 'Treatment of adult acute lymphoblastic leukemia (ALL): long-term follow-up of the GIMEMA ALL 0288 randomized study', *Blood*, vol. 99, no. 3, pp. 863-71.

Arico, M., Valsecchi, M. G., Camitta, B., Schrappe, M., Chessells, J., Baruchel, A., Gaynon, P., Silverman, L., Janka-Schaub, G., Kamps, W., Pui, C. H. & Masera, G.

2000, 'Outcome of treatment in children with Philadelphia chromosome-positive acute lymphoblastic leukemia', *N Engl J Med*, vol. 342, no. 14, pp. 998-1006.

Arico, M., Valsecchi, M. G., Rizzari, C., Barisone, E., Biondi, A., Casale, F., Locatelli, F., Lo Nigro, L., Luciani, M., Messina, C., Micalizzi, C., Parasole, R., Pession, A., Santoro, N., Testi, A. M., Silvestri, D., Basso, G., Masera, G. & Conter, V. 2008, 'Long-term results of the AIEOP-ALL-95 Trial for Childhood Acute Lymphoblastic Leukemia: insight on the prognostic value of DNA index in the framework of Berlin-Frankfurt-Muenster based chemotherapy', *J Clin Oncol*, vol. 26, no. 2, pp. 283-9.

Arnott, J. A., Zhang, X., Sanjay, A., Owen, T. A., Smock, S. L., Rehman, S., DeLong, W. G., Safadi, F. F. & Popoff, S. N. 2008, 'Molecular requirements for induction of CTGF expression by TGF-beta1 in primary osteoblasts', *Bone*, vol. 42, no. 5, pp. 871-85.

Aspland, S. E., Bendall, H. H. & Murre, C. 2001, 'The role of E2A-PBX1 in leukemogenesis', *Oncogene*, vol. 20, no. 40, pp. 5708-17.

Australian Bureau of Statistics 2006, *Causes of Death, Australia (2004)*, 3303.0, Australian Bureau of Statistics, Canberra.

Ayton, P. M. & Cleary, M. L. 2001, 'Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins', *Oncogene*, vol. 20, no. 40, pp. 5695-707.

Bailey, T. L. & Elkan, C. 1994, 'Fitting a mixture model by expectation maximization to discover motifs in biopolymers', *Proc Int Conf Intell Syst Mol Biol*, vol. 2, pp. 28-36.

Bain, G., Romanow, W. J., Albers, K., Havran, W. L. & Murre, C. 1999, 'Positive and negative regulation of V(D)J recombination by the E2A proteins', *J Exp Med*, vol. 189, no. 2, pp. 289-300.

Ball, D. K., Moussad, E. E., Rageh, M. A., Kemper, S. A. & Brigstock, D. R. 2003a, 'Establishment of a recombinant expression system for connective tissue growth factor (CTGF) that models CTGF processing in utero', *Reproduction*, vol. 125, no. 2, pp. 271-84.

Ball, D. K., Rachfal, A. W., Kemper, S. A. & Brigstock, D. R. 2003b, 'The heparin-binding 10 kDa fragment of connective tissue growth factor (CTGF) containing module 4 alone stimulates cell adhesion', *J Endocrinol*, vol. 176, no. 2, pp. R1-7.

Ball, D. K., Surveyor, G. A., Diehl, J. R., Steffen, C. L., Uzumcu, M., Mirando, M. A. & Brigstock, D. R. 1998, 'Characterization of 16- to 20-kilodalton (kDa) connective tissue growth factors (CTGFs) and demonstration of proteolytic activity for 38-kDa CTGF in pig uterine luminal flushings', *Biol Reprod*, vol. 59, no. 4, pp. 828-35.

Barbolina, M. V., Adley, B. P., Kelly, D. L., Shepard, J., Fought, A. J., Scholtens, D., Penzes, P., Shea, L. D. & Stack, M. S. 2009, 'Downregulation of connective tissue growth factor by three-dimensional matrix enhances ovarian carcinoma cell invasion', *Int J Cancer*, vol. 125, no. 4, pp. 816-25.

Barth, P. J. & Westhoff, C. C. 2007, 'CD34+ fibrocytes: morphology, histogenesis and function', *Curr Stem Cell Res Ther*, vol. 2, no. 3, pp. 221-7.

Beesley, A. H., Palmer, M. L., Ford, J., Weller, R. E., Cummings, A. J., Freitas, J. R., Firth, M. J., Perera, K. U., de Klerk, N. H. & Kees, U. R. 2006, 'Authenticity and drug resistance in a panel of acute lymphoblastic leukaemia cell lines', *Br J Cancer*, vol. 95, no. 11, pp. 1537-44.

Bene, M. C. 2005, 'Immunophenotyping of acute leukaemias', *Immunol Lett*, vol. 98, no. 1, pp. 9-21.

Bernstein, B. E., Meissner, A. & Lander, E. S. 2007, 'The mammalian epigenome', *Cell*, vol. 128, no. 4, pp. 669-81.

Bhatia, S. 2004, 'Influence of race and socioeconomic status on outcome of children treated for childhood acute lymphoblastic leukemia', *Curr Opin Pediatr*, vol. 16, no. 1, pp. 9-14.

Bjourson, A. J. & Cooper, J. E. 1992, 'Band-stab PCR: a simple technique for the purification of individual PCR products', *Nucleic Acids Res*, vol. 20, no. 17, p. 4675.

Blackledge, N. P., Zhou, J. C., Tolstorukov, M. Y., Farcas, A. M., Park, P. J. & Klose, R. J. 2010, 'CpG islands recruit a histone H3 lysine 36 demethylase', *Mol Cell*, vol. 38, no. 2, pp. 179-90.

Blesa, S., Vernia, S., Garcia-Garcia, A. B., Martinez-Hervas, S., Ivorra, C., Gonzalez-Albert, V., Ascaso, J. F., Martin-Escudero, J. C., Real, J. T., Carmena, R., Casado, M. & Chaves, F. J. 2008, 'A new PCSK9 gene promoter variant affects gene expression and causes autosomal dominant hypercholesterolemia', *J Clin Endocrinol Metab*, vol. 93, no. 9, pp. 3577-83.

Blom, B. & Spits, H. 2006, 'Development of human lymphoid cells', *Annu Rev Immunol*, vol. 24, pp. 287-320.

Boag, J. M., Beesley, A. H., Firth, M. J., Freitas, J. R., Ford, J., Brigstock, D. R., de Klerk, N. H. & Kees, U. R. 2007, 'High expression of connective tissue growth factor in pre-B acute lymphoblastic leukaemia', *Br J Haematol*, vol. 138, no. 6, pp. 740-8.

Boag, J. M., Beesley, A. H., Firth, M. J., Freitas, J. R., Ford, J., Hoffmann, K., Cummings, A. J., de Klerk, N. H. & Kees, U. R. 2006, 'Altered glucose metabolism in childhood pre-B acute lymphoblastic leukaemia', *Leukemia*, vol. 20, no. 10, pp. 1731-7.

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J. & Lengauer, T. 2005, 'BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing', *Bioinformatics*, vol. 21, no. 21, pp. 4067-8.

Boidot, R., Vegran, F., Jacob, D., Chevrier, S., Cadouot, M., Feron, O., Solary, E. & Lizard-Nacol, S. 2010, 'The transcription factor GATA-1 is overexpressed in breast carcinomas and contributes to survivin upregulation via a promoter polymorphism', *Oncogene*, vol. 29, no. 17, pp. 2577-84.

Bonifer, C. & Bowen, D. T. 2010, 'Epigenetic mechanisms regulating normal and malignant haematopoiesis: new therapeutic targets for clinical medicine', *Expert Rev Mol Med*, vol. 12, p. e6.

Bourquin, J. P. & Izraeli, S. 2010, 'Where can biology of childhood ALL be attacked by new compounds?' *Cancer Treat Rev*, vol. 36, no. 4, pp. 298-306.

Bradford, M. M. 1976, 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Anal Biochem*, vol. 72, pp. 248-54.

Bradham, D. M., Igarashi, A., Potter, R. L. & Grotendorst, G. R. 1991, 'Connective tissue growth factor: a cysteine-rich mitogen secreted by human vascular endothelial cells is related to the SRC-induced immediate early gene product CEF-10', *J Cell Biol*, vol. 114, no. 6, pp. 1285-94.

Brigstock, D. R. 2003, 'The CCN family: a new stimulus package', *J Endocrinol*, vol. 178, no. 2, pp. 169-75.

Brigstock, D. R. 2010, 'Connective tissue growth factor (CCN2, CTGF) and organ fibrosis: lessons from transgenic animals', *J Cell Commun Signal*, vol. 4, no. 1, pp. 1-4.

Brigstock, D. R., Steffen, C. L., Kim, G. Y., Vegunta, R. K., Diehl, J. R. & Harding, P. A. 1997, 'Purification and characterization of novel heparin-binding growth factors in uterine secretory fluids. Identification as heparin-regulated Mr 10,000 forms of connective tissue growth factor', *J Biol Chem*, vol. 272, no. 32, pp. 20275-82.

Brothman, A. R., Swanson, G., Maxwell, T. M., Cui, J., Murphy, K. J., Herrick, J., Speights, V. O., Isaac, J. & Rohr, L. R. 2005, 'Global hypomethylation is common in prostate cancer cells: a quantitative predictor for clinical outcome?' *Cancer Genet Cytogenet*, vol. 156, no. 1, pp. 31-6.

Brown, P., Levis, M., Shurtleff, S., Campana, D., Downing, J. & Small, D. 2005, 'FLT3 inhibition selectively kills childhood acute lymphoblastic leukemia cells with high levels of FLT3 expression', *Blood*, vol. 105, no. 2, pp. 812-20.

Bruggemann, M., Schrauder, A., Raff, T., Pfeifer, H., Dworzak, M., Ottmann, O. G., Asnafi, V., Baruchel, A., Bassan, R., Benoit, Y., Biondi, A., Cave, H., Dombret, H., Fielding, A. K., Foa, R., Gokbuget, N., Goldstone, A. H., Goulden, N., Henze, G., Hoelzer, D., Janka-Schaub, G. E., Macintyre, E. A., Pieters, R., Rambaldi, A., Ribera, J. M., Schmiegelow, K., Spinelli, O., Stary, J., von Stackelberg, A., Kneba, M., Schrappe, M. & van Dongen, J. J. 2010, 'Standardized MRD quantification in European ALL trials: proceedings of the Second International Symposium on MRD assessment in Kiel, Germany, 18-20 September 2008', *Leukemia*, vol. 24, no. 3, pp. 521-35.

Buffler, P. A., Kwan, M. L., Reynolds, P. & Urayama, K. Y. 2005, 'Environmental and genetic risk factors for childhood leukemia: appraising the evidence', *Cancer Invest*, vol. 23, no. 1, pp. 60-75.

Cadioux, B., Ching, T. T., VandenBerg, S. R. & Costello, J. F. 2006, 'Genome-wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation', *Cancer Res*, vol. 66, no. 17, pp. 8469-76.

Cervello, M., Giannitrapani, L., Labbozzetta, M., Notarbartolo, M., D'Alessandro, N., Lampiasi, N., Azzolina, A. & Montalto, G. 2004, 'Expression of WISPs and of their novel alternative variants in human hepatocellular carcinoma cells', *Ann N Y Acad Sci*, vol. 1028, pp. 432-9.

Chaqour, B. & Goppelt-Struebe, M. 2006, 'Mechanical regulation of the Cyr61/CCN1 and CTGF/CCN2 proteins', *FEBS J*, vol. 273, no. 16, pp. 3639-49.

Chen, C. C. & Lau, L. F. 2009, 'Functions and mechanisms of action of CCN matricellular proteins', *Int J Biochem Cell Biol*, vol. 41, no. 4, pp. 771-83.

Chen, N., Chen, C. C. & Lau, L. F. 2000, 'Adhesion of human skin fibroblasts to Cyr61 is mediated through integrin alpha 6beta 1 and cell surface heparan sulfate proteoglycans', *J Biol Chem*, vol. 275, no. 32, pp. 24953-61.

Chen, Y., Blom, I. E., Sa, S., Goldschmeding, R., Abraham, D. J. & Leask, A. 2002, 'CTGF expression in mesangial cells: involvement of SMADs, MAP kinase, and PKC', *Kidney Int*, vol. 62, no. 4, pp. 1149-59.

Chen, Y., Segarini, P., Raoufi, F., Bradham, D. & Leask, A. 2001, 'Connective tissue growth factor is secreted through the Golgi and is degraded in the endosome', *Exp Cell Res*, vol. 271, no. 1, pp. 109-17.

Chi, P., Allis, C. D. & Wang, G. G. 2010, 'Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers', *Nat Rev Cancer*, vol. 10, no. 7, pp. 457-69.

Chiba, T., Yokosuka, O., Arai, M., Tada, M., Fukai, K., Imazeki, F., Kato, M., Seki, N. & Saisho, H. 2004, 'Identification of genes up-regulated by histone deacetylase

inhibition with cDNA microarray and exploration of epigenetic alterations on hepatoma cells', *J Hepatol*, vol. 41, no. 3, pp. 436-45.

Chiba, T., Yokosuka, O., Fukai, K., Hirasawa, Y., Tada, M., Mikata, R., Imazeki, F., Taniguchi, H., Iwama, A., Miyazaki, M., Ochiai, T. & Saisho, H. 2005, 'Identification and investigation of methylated genes in hepatoma', *Eur J Cancer*, vol. 41, no. 8, pp. 1185-94.

Chien, W., Yin, D., Gui, D., Mori, A., Frank, J. M., Said, J., Kusuanco, D., Marchevsky, A., McKenna, R. & Koeffler, H. P. 2006, 'Suppression of cell proliferation and signaling transduction by connective tissue growth factor in non-small cell lung cancer cells', *Mol Cancer Res*, vol. 4, no. 8, pp. 591-8.

Choi, J., Hwang, Y. K., Sung, K. W., Kim, D. H., Yoo, K. H., Jung, H. L. & Koo, H. H. 2006, 'Aven overexpression: association with poor prognosis in childhood acute lymphoblastic leukemia', *Leuk Res*, vol. 30, no. 8, pp. 1019-25.

Chouaib, S., Kieda, C., Benlalam, H., Noman, M. Z., Mami-Chouaib, F. & Ruegg, C. 2011, 'Endothelial cells as key determinants of the tumor microenvironment: interaction with tumor cells, extracellular matrix and immune killer cells', *Crit Rev Immunol*, vol. 30, no. 6, pp. 529-45.

Chu, C. Y., Chang, C. C., Prakash, E. & Kuo, M. L. 2008, 'Connective tissue growth factor (CTGF) and cancer progression', *J Biomed Sci*, vol. 15, no. 6, pp. 675-85.

Cicha, I. & Goppelt-Struebe, M. 2009, 'Connective tissue growth factor: context-dependent functions and mechanisms of regulation', *Biofactors*, vol. 35, no. 2, pp. 200-8.

Cobb, B. S. & Smale, S. T. 2005, 'Ikaros-family proteins: in search of molecular functions during lymphocyte development', *Curr Top Microbiol Immunol*, vol. 290, pp. 29-47.

Conter, V., Arico, M., Basso, G., Biondi, A., Barisone, E., Messina, C., Parasole, R., De Rossi, G., Locatelli, F., Pession, A., Santoro, N., Micalizzi, C., Citterio, M., Rizzari, C., Silvestri, D., Rondelli, R., Lo Nigro, L., Ziino, O., Testi, A. M., Masera, G. & Valsecchi, M. G. 2010, 'Long-term results of the Italian Association of Pediatric Hematology and Oncology (AIEOP) Studies 82, 87, 88, 91 and 95 for childhood acute lymphoblastic leukemia', *Leukemia*, vol. 24, no. 2, pp. 255-64.

Corfe, S. & Paige, C. 2009, 'Development of B Lymphocytes', in *Molecular Basis of Hematopoiesis*, eds A. Wickrema & B. Kee, Springer, New York, pp. 173-199.

Crisan, M., Yap, S., Casteilla, L., Chen, C. W., Corselli, M., Park, T. S., Andriolo, G., Sun, B., Zheng, B., Zhang, L., Norotte, C., Teng, P. N., Traas, J., Schugar, R., Deasy, B. M., Badylak, S., Buhning, H. J., Giacobino, J. P., Lazzari, L., Huard, J. &

Peault, B. 2008, 'A perivascular origin for mesenchymal stem cells in multiple human organs', *Cell Stem Cell*, vol. 3, no. 3, pp. 301-13.

De Angioletti, M., Lacerra, G., Sabato, V. & Carestia, C. 2004, 'Beta+45 G --> C: a novel silent beta-thalassaemia mutation, the first in the Kozak sequence', *Br J Haematol*, vol. 124, no. 2, pp. 224-31.

de Capoa, A., Musolino, A., Della Rosa, S., Caiafa, P., Mariani, L., Del Nonno, F., Vocaturo, A., Donnorso, R. P., Niveleau, A. & Grappelli, C. 2003, 'DNA demethylation is directly related to tumour progression: evidence in normal, pre-malignant and malignant cells from uterine cervix samples', *Oncol Rep*, vol. 10, no. 3, pp. 545-9.

Denoeud, F., Kapranov, P., Ucla, C., Frankish, A., Castelo, R., Drenkow, J., Lagarde, J., Alioto, T., Manzano, C., Chrast, J., Dike, S., Wyss, C., Henrichsen, C. N., Holroyd, N., Dickson, M. C., Taylor, R., Hance, Z., Foissac, S., Myers, R. M., Rogers, J., Hubbard, T., Harrow, J., Guigo, R., Gingeras, T. R., Antonarakis, S. E. & Reymond, A. 2007, 'Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions', *Genome Res*, vol. 17, no. 6, pp. 746-59.

Dhar, A. & Ray, A. 2010, 'The CCN family proteins in carcinogenesis', *Exp Oncol*, vol. 32, no. 1, pp. 2-9.

Dias, S., Silva, H., Jr., Cumano, A. & Vieira, P. 2005, 'Interleukin-7 is necessary to maintain the B cell potential in common lymphoid progenitors', *J Exp Med*, vol. 201, no. 6, pp. 971-9.

Dowling, E., Yabroff, K. R., Mariotto, A., McNeel, T., Zeruto, C. & Buckman, D. 2010, 'Burden of illness in adult survivors of childhood cancers: findings from a population-based national sample', *Cancer*, vol. 116, no. 15, pp. 3712-21.

Duisters, R. F., Tijssen, A. J., Schroen, B., Leenders, J. J., Lentink, V., van der Made, I., Herias, V., van Leeuwen, R. E., Schellings, M. W., Barenbrug, P., Maessen, J. G., Heymans, S., Pinto, Y. M. & Creemers, E. E. 2009, 'miR-133 and miR-30 regulate connective tissue growth factor: implications for a role of microRNAs in myocardial matrix remodeling', *Circ Res*, vol. 104, no. 2, pp. 170-8.

Ebinger, M., Senf, L., Wachowski, O. & Scheurlen, W. 2004, 'Promoter methylation pattern of caspase-8, P16INK4A, MGMT, TIMP-3, and E-cadherin in medulloblastoma', *Pathol Oncol Res*, vol. 10, no. 1, pp. 17-21.

Eguchi, T., Kubota, S., Kawata, K., Mukudai, Y., Ohgawara, T., Miyazono, K., Nakao, K., Kondo, S. & Takigawa, M. 2007, 'Different transcriptional strategies for *ccn2/ctgf* gene induction between human chondrocytic and breast cancer cell lines', *Biochimie*, vol. 89, no. 3, pp. 278-88.

- Eguchi, T., Kubota, S., Kawata, K., Mukudai, Y., Uehara, J., Ohgawara, T., Ibaragi, S., Sasaki, A., Kuboki, T. & Takigawa, M. 2008, 'Novel transcription-factor-like function of human matrix metalloproteinase 3 regulating the CTGF/CCN2 gene', *Mol Cell Biol*, vol. 28, no. 7, pp. 2391-413.
- Ehninger, A. & Trumpp, A. 2011, 'The bone marrow stem cell niche grows up: mesenchymal stem cells and macrophages move in', *J Exp Med*, vol. 208, no. 3, pp. 421-8.
- Ehrlich, M. 2009, 'DNA hypomethylation in cancer cells', *Epigenomics*, vol. 1, no. 2, pp. 239-259.
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. 1998, 'Cluster analysis and display of genome-wide expression patterns', *Proc Natl Acad Sci U S A*, vol. 95, no. 25, pp. 14863-8.
- Elmlinger, M. W., Wimmer, K., Biemer, E., Blum, W. F., Ranke, M. B. & Dannecker, G. E. 1996, 'Insulin-like growth factor binding protein 2 is differentially expressed in leukaemic B- and T-cell lines', *Growth Regul*, vol. 6, no. 3, pp. 152-7.
- Ernst, A., Campos, B., Meier, J., Devens, F., Liesenberg, F., Wolter, M., Reifemberger, G., Herold-Mende, C., Lichter, P. & Radlwimmer, B. 2010, 'De-repression of CTGF via the miR-17-92 cluster upon differentiation of human glioblastoma spheroid cultures', *Oncogene*, vol. 29, no. 23, pp. 3411-22.
- Farre, D., Roset, R., Huerta, M., Adsuara, J. E., Rosello, L., Alba, M. M. & Messeguer, X. 2003, 'Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN', *Nucleic Acids Res*, vol. 31, no. 13, pp. 3651-3.
- Faull, R. J. 2005, 'Bad and good growth factors in the peritoneal cavity', *Nephrology*, vol. 10, no. 3, pp. 234-9.
- Feinberg, A. P. & Vogelstein, B. 1983a, 'Hypomethylation distinguishes genes of some human cancers from their normal counterparts', *Nature*, vol. 301, no. 5895, pp. 89-92.
- Feinberg, A. P. & Vogelstein, B. 1983b, 'Hypomethylation of ras oncogenes in primary human cancers', *Biochem Biophys Res Commun*, vol. 111, no. 1, pp. 47-54.
- Feng, Q. & Zhang, Y. 2001, 'The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes', *Genes Dev*, vol. 15, no. 7, pp. 827-32.
- Fischer, M., Schwieger, M., Horn, S., Niebuhr, B., Ford, A., Roscher, S., Bergholz, U., Greaves, M., Lohler, J. & Stocking, C. 2005, 'Defining the oncogenic function of

the TEL/AML1 (ETV6/RUNX1) fusion protein in a mouse model', *Oncogene*, vol. 24, no. 51, pp. 7579-91.

Fonseca, C., Lindahl, G. E., Ponticos, M., Sestini, P., Renzoni, E. A., Holmes, A. M., Spagnolo, P., Pantelidis, P., Leoni, P., McHugh, N., Stock, C. J., Shi-Wen, X., Denton, C. P., Black, C. M., Welsh, K. I., du Bois, R. M. & Abraham, D. J. 2007, 'A polymorphism in the CTGF promoter region associated with systemic sclerosis', *N Engl J Med*, vol. 357, no. 12, pp. 1210-20.

Frazier, K. S. & Grotendorst, G. R. 1997, 'Expression of connective tissue growth factor mRNA in the fibrous stroma of mammary tumors', *Int J Biochem Cell Biol*, vol. 29, no. 1, pp. 153-61.

Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. 2009, 'Most mammalian mRNAs are conserved targets of microRNAs', *Genome Res*, vol. 19, no. 1, pp. 92-105.

Fu, M., Zhang, J., Zhu, X., Myles, D. E., Willson, T. M., Liu, X. & Chen, Y. E. 2001, 'Peroxisome proliferator-activated receptor gamma inhibits transforming growth factor beta-induced connective tissue growth factor expression in human aortic smooth muscle cells by interfering with Smad3', *J Biol Chem*, vol. 276, no. 49, pp. 45888-94.

Fu, Y. & Weng, Z. 2004, 'Improvement of TRANSFAC matrices using multiple local alignment of transcription factor binding site sequences', *Conf Proc IEEE Eng Med Biol Soc*, vol. 4, pp. 2856-9.

Gala, J. L., Vermynen, C., Cornu, G., Ferrant, A., Michaux, J. L., Philippe, M. & Martiat, P. 1994, 'High expression of bcl-2 is the rule in acute lymphoblastic leukemia, except in Burkitt subtype at presentation, and is not correlated with the prognosis', *Ann Hematol*, vol. 69, no. 1, pp. 17-24.

Gama-Sosa, M. A., Slagel, V. A., Trewyn, R. W., Oxenhandler, R., Kuo, K. C., Gehrke, C. W. & Ehrlich, M. 1983, 'The 5-methylcytosine content of DNA from human tumors', *Nucleic Acids Res*, vol. 11, no. 19, pp. 6883-94.

Gao, R. & Brigstock, D. R. 2003, 'Low density lipoprotein receptor-related protein (LRP) is a heparin-dependent adhesion receptor for connective tissue growth factor (CTGF) in rat activated hepatic stellate cells', *Hepatol Res*, vol. 27, no. 3, pp. 214-220.

Gao, R. & Brigstock, D. R. 2004, 'Connective tissue growth factor (CCN2) induces adhesion of rat activated hepatic stellate cells by binding of its C-terminal domain to integrin alpha(v)beta(3) and heparan sulfate proteoglycan', *J Biol Chem*, vol. 279, no. 10, pp. 8848-55.

Gao, R. & Brigstock, D. R. 2006, 'A novel integrin alpha5beta1 binding domain in module 4 of connective tissue growth factor (CCN2/CTGF) promotes adhesion and migration of activated pancreatic stellate cells', *Gut*, vol. 55, no. 6, pp. 856-62.

Gaynon, P. S. 2005, 'Childhood acute lymphoblastic leukaemia and relapse', *Br J Haematol*, vol. 131, no. 5, pp. 579-87.

Gaynon, P. S., Angiolillo, A. L., Carroll, W. L., Nachman, J. B., Trigg, M. E., Sather, H. N., Hunger, S. P. & Devidas, M. 2010, 'Long-term results of the children's cancer group studies for childhood acute lymphoblastic leukemia 1983-2002: a Children's Oncology Group Report', *Leukemia*, vol. 24, no. 2, pp. 285-97.

Gaynon, P. S., Trigg, M. E., Heerema, N. A., Sensel, M. G., Sather, H. N., Hammond, G. D. & Bleyer, W. A. 2000, 'Children's Cancer Group trials in childhood acute lymphoblastic leukemia: 1983-1995', *Leukemia*, vol. 14, no. 12, pp. 2223-33.

Geenen, M. M., Bakker, P. J., Kremer, L. C., Kastelein, J. J. & van Leeuwen, F. E. 2010, 'Increased prevalence of risk factors for cardiovascular disease in long-term survivors of acute lymphoblastic leukemia and Wilms tumor treated with radiotherapy', *Pediatr Blood Cancer*, vol. 55, no. 4, pp. 690-7.

Georgopoulos, K., Bigby, M., Wang, J. H., Molnar, A., Wu, P., Winandy, S. & Sharpe, A. 1994, 'The Ikaros gene is required for the development of all lymphoid lineages', *Cell*, vol. 79, no. 1, pp. 143-56.

Gourh, P., Mayes, M. D. & Arnett, F. C. 2008, 'CTGF polymorphism associated with systemic sclerosis', *N Engl J Med*, vol. 358, no. 3, pp. 308-9; author reply 309.

Granel, B., Argiro, L., Hachulla, E., Fajardy, I., Weiller, P. J., Durand, J. M., Frances, Y., Dombey, A. M., Marquet, S., Lesavre, N., Disdier, P., Bernard, F., Hatron, P. Y. & Chevillard, C. 2010, 'Association between a CTGF gene polymorphism and systemic sclerosis in a French population', *J Rheumatol*, vol. 37, no. 2, pp. 351-8.

Greaves, M. 2009, 'Darwin and evolutionary tales in leukemia. The Ham-Wasserman Lecture', *Hematology Am Soc Hematol Educ Program*, pp. 3-12.

Gressner, O. A. & Gressner, A. M. 2008, 'Connective tissue growth factor: a fibrogenic master switch in fibrotic liver diseases', *Liver Int*, vol. 28, no. 8, pp. 1065-79.

Grotendorst, G. R. 1997, 'Connective tissue growth factor: a mediator of TGF-beta action on fibroblasts', *Cytokine Growth Factor Rev*, vol. 8, no. 3, pp. 171-9.

Grotendorst, G. R., Okochi, H. & Hayashi, N. 1996, 'A novel transforming growth factor beta response element controls the expression of the connective tissue growth factor gene', *Cell Growth Differ*, vol. 7, no. 4, pp. 469-80.

Guenther, M. G., Lawton, L. N., Rozovskaia, T., Frampton, G. M., Levine, S. S., Volkert, T. L., Croce, C. M., Nakamura, T., Canaani, E. & Young, R. A. 2008, 'Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia', *Genes Dev*, vol. 22, no. 24, pp. 3403-8.

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. 2007, 'Quantifying similarity between motifs', *Genome Biol*, vol. 8, no. 2, p. R24.

Gurney, J. G., Severson, R. K., Davis, S. & Robison, L. L. 1995, 'Incidence of cancer in children in the United States. Sex-, race-, and 1-year age-specific rates by histologic type', *Cancer*, vol. 75, no. 8, pp. 2186-95.

Hallegger, M., Llorian, M. & Smith, C. W. 2010, 'Alternative splicing: global insights', *FEBS J*, vol. 277, no. 4, pp. 856-66.

Hanada, M., Delia, D., Aiello, A., Stadtmauer, E. & Reed, J. C. 1993, 'bcl-2 gene hypomethylation and high-level expression in B-cell chronic lymphocytic leukemia', *Blood*, vol. 82, no. 6, pp. 1820-8.

Harrison, C. J., Moorman, A. V., Broadfield, Z. J., Cheung, K. L., Harris, R. L., Reza Jalali, G., Robinson, H. M., Barber, K. E., Richards, S. M., Mitchell, C. D., Eden, T. O., Hann, I. M., Hill, F. G., Kinsey, S. E., Gibson, B. E., Lilleyman, J., Vora, A., Goldstone, A. H., Franklin, I. M., Durrant, J. & Martineau, M. 2004, 'Three distinct subgroups of hypodiploidy in acute lymphoblastic leukaemia', *Br J Haematol*, vol. 125, no. 5, pp. 552-9.

Harvey, R. C., Mullighan, C. G., Wang, X., Dobbin, K. K., Davidson, G. S., Bedrick, E. J., Chen, I. M., Atlas, S. R., Kang, H., Ar, K., Wilson, C. S., Wharton, W., Murphy, M., Devidas, M., Carroll, A. J., Borowitz, M. J., Bowman, W. P., Downing, J. R., Relling, M., Yang, J., Bhojwani, D., Carroll, W. L., Camitta, B., Reaman, G. H., Smith, M., Hunger, S. P. & Willman, C. L. 2010, 'Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome', *Blood*, vol. 116, no. 23, pp. 4874-84.

Hashimoto, G., Inoki, I., Fujii, Y., Aoki, T., Ikeda, E. & Okada, Y. 2002, 'Matrix metalloproteinases cleave connective tissue growth factor and reactivate angiogenic activity of vascular endothelial growth factor 165', *J Biol Chem*, vol. 277, no. 39, pp. 36288-95.

Hatakeyama, M., Tsudo, M., Minamoto, S., Kono, T., Doi, T., Miyata, T., Miyasaka, M. & Taniguchi, T. 1989, 'Interleukin-2 receptor beta chain gene: generation of three receptor forms by cloned human alpha and beta chain cDNA's', *Science*, vol. 244, no. 4904, pp. 551-6.

He, S., Jin, M. L., Worpel, V. & Hinton, D. R. 2003, 'A role for connective tissue growth factor in the pathogenesis of choroidal neovascularization', *Arch Ophthalmol*, vol. 121, no. 9, pp. 1283-8.

Heerema, N. A., Sather, H. N., Sensel, M. G., Zhang, T., Hutchinson, R. J., Nachman, J. B., Lange, B. J., Steinherz, P. G., Bostrom, B. C., Reaman, G. H., Gaynon, P. S. & Uckun, F. M. 2000, 'Prognostic impact of trisomies of chromosomes 10, 17, and 5 among children with acute lymphoblastic leukemia and high hyperdiploidy (> 50 chromosomes)', *J Clin Oncol*, vol. 18, no. 9, pp. 1876-87.

Higgins, D. F., Biju, M. P., Akai, Y., Wutz, A., Johnson, R. S. & Haase, V. H. 2004, 'Hypoxic induction of Ctgf is directly mediated by Hif-1', *Am J Physiol Renal Physiol*, vol. 287, no. 6, pp. F1223-32.

Hirschfeld, M., zur Hausen, A., Bettendorf, H., Jager, M. & Stickeler, E. 2009, 'Alternative splicing of Cyr61 is regulated by hypoxia and significantly changed in breast cancer', *Cancer Res*, vol. 69, no. 5, pp. 2082-90.

Hjalgrim, L. L., Madsen, H. O., Melbye, M., Jorgensen, P., Christiansen, M., Andersen, M. T., Pallisgaard, N., Hokland, P., Clausen, N., Ryder, L. P., Schmiegelow, K. & Hjalgrim, H. 2002, 'Presence of clone-specific markers at birth in children with acute lymphoblastic leukaemia', *Br J Cancer*, vol. 87, no. 9, pp. 994-9.

Ho, J. W., Bishop, E., Kharchenko, P. V., Negre, N., White, K. P. & Park, P. J. 2011, 'ChIP-chip versus ChIP-seq: Lessons for experimental design and data analysis', *BMC Genomics*, vol. 12, no. 1, p. 134.

Hoffmann, K., Firth, M. J., Beesley, A. H., de Klerk, N. H. & Kees, U. R. 2006, 'Translating microarray data for diagnostic testing in childhood leukaemia', *BMC Cancer*, vol. 6, p. 229-240.

Hoffmann, K., Firth, M. J., Freitas, J. R., de Klerk, N. H. & Kees, U. R. 2005, 'Gene expression levels in small specimens from patients detected using oligonucleotide arrays', *Mol Biotechnol*, vol. 29, no. 1, pp. 31-8.

Holbourn, K. P., Acharya, K. R. & Perbal, B. 2008, 'The CCN family of proteins: structure-function relationships', *Trends Biochem Sci*, vol. 33, no. 10, pp. 461-73.

Holliday, R. & Pugh, J. E. 1975, 'DNA modification mechanisms and gene activity during development', *Science*, vol. 187, no. 4173, pp. 226-32.

Holmes, A., Abraham, D. J., Chen, Y., Denton, C., Shi-wen, X., Black, C. M. & Leask, A. 2003, 'Constitutive connective tissue growth factor expression in scleroderma fibroblasts is dependent on Sp1', *J Biol Chem*, vol. 278, no. 43, pp. 41728-33.

- Holmes, A., Abraham, D. J., Sa, S., Shiwen, X., Black, C. M. & Leask, A. 2001, 'CTGF and SMADs, maintenance of scleroderma phenotype is independent of SMAD signaling', *J Biol Chem*, vol. 276, no. 14, pp. 10594-601.
- Hong, D., Gupta, R., Ancliff, P., Atzberger, A., Brown, J., Soneji, S., Green, J., Colman, S., Piacibello, W., Buckle, V., Tsuzuki, S., Greaves, M. & Enver, T. 2008, 'Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia', *Science*, vol. 319, no. 5861, pp. 336-9.
- Hong, K. H., Yoo, S. A., Kang, S. S., Choi, J. J., Kim, W. U. & Cho, C. S. 2006, 'Hypoxia induces expression of connective tissue growth factor in scleroderma skin fibroblasts', *Clin Exp Immunol*, vol. 146, no. 2, pp. 362-70.
- Honorio, S., Agathangelou, A., Wernert, N., Rothe, M., Maher, E. R. & Latif, F. 2003, 'Frequent epigenetic inactivation of the RASSF1A tumour suppressor gene in testicular tumours and distinct methylation profiles of seminoma and nonseminoma testicular germ cell tumours', *Oncogene*, vol. 22, no. 3, pp. 461-6.
- Hoshijima, M., Hattori, T., Inoue, M., Araki, D., Hanagata, H., Miyauchi, A. & Takigawa, M. 2006, 'CT domain of CCN2/CTGF directly interacts with fibronectin and enhances cell adhesion of chondrocytes through integrin alpha5beta1', *FEBS Lett*, vol. 580, no. 5, pp. 1376-82.
- Hosokawa, Y., Maeda, Y. & Seto, M. 1999, 'Human Helios, an Ikaros-related zinc finger DNA binding protein: cDNA cloning and tissue expression pattern', *Immunogenetics*, vol. 50, no. 1-2, pp. 106-8.
- Houseley, J. & Tollervey, D. 2009, 'The many pathways of RNA degradation', *Cell*, vol. 136, no. 4, pp. 763-76.
- Huang, B. L., Dornbach, L. M. & Lyons, K. M. 2007, 'The 5' untranslated regions (UTRs) of CCN1, CCN2, and CCN4 exhibit cryptic promoter activity', *J Cell Commun Signal*, vol. 1, no. 1, pp. 17-32.
- Huang, E. J. & Reichardt, L. F. 2003, 'Trk receptors: roles in neuronal signal transduction', *Annu Rev Biochem*, vol. 72, pp. 609-42.
- Iacobucci, I., Lonetti, A., Messa, F., Cilloni, D., Arruga, F., Ottaviani, E., Paolini, S., Papayannidis, C., Piccaluga, P. P., Giannoulia, P., Soverini, S., Amabile, M., Poerio, A., Saglio, G., Pane, F., Berton, G., Baruzzi, A., Vitale, A., Chiaretti, S., Perini, G., Foa, R., Baccarani, M. & Martinelli, G. 2008, 'Expression of spliced oncogenic Ikaros isoforms in Philadelphia-positive acute lymphoblastic leukemia patients treated with tyrosine kinase inhibitors: implications for a new mechanism of resistance', *Blood*, vol. 112, no. 9, pp. 3847-55.
- Igarashi, A., Nashiro, K., Kikuchi, K., Sato, S., Ihn, H., Grotendorst, G. R. & Takehara, K. 1995, 'Significant correlation between connective tissue growth factor

gene expression and skin sclerosis in tissue sections from patients with systemic sclerosis', *J Invest Dermatol*, vol. 105, no. 2, pp. 280-4.

Igarashi, A., Okochi, H., Bradham, D. M. & Grotendorst, G. R. 1993, 'Regulation of connective tissue growth factor gene expression in human skin fibroblasts and during wound repair', *Mol Biol Cell*, vol. 4, no. 6, pp. 637-45.

Inoki, I., Shiomi, T., Hashimoto, G., Enomoto, H., Nakamura, H., Makino, K., Ikeda, E., Takata, S., Kobayashi, K. & Okada, Y. 2002, 'Connective tissue growth factor binds vascular endothelial growth factor (VEGF) and inhibits VEGF-induced angiogenesis', *FASEB J*, vol. 16, no. 2, pp. 219-21.

Ishida, Y., Sakamoto, N., Kamibeppu, K., Kakee, N., Iwai, T., Ozono, S., Maeda, N., Okamura, J., Asami, K., Inada, H., Honda, M. & Horibe, K. 2010, 'Late effects and quality of life of childhood cancer survivors: Part 2. Impact of radiotherapy', *Int J Hematol*, vol. 92, no. 1, pp. 95-104.

Ito, T. K., Ishii, G., Chiba, H. & Ochiai, A. 2007, 'The VEGF angiogenic switch of fibroblasts is regulated by MMP-7 from cancer cells', *Oncogene*, vol. 26, no. 51, pp. 7194-203.

Ivkovic, S., Yoon, B. S., Popoff, S. N., Safadi, F. F., Libuda, D. E., Stephenson, R. C., Daluiski, A. & Lyons, K. M. 2003, 'Connective tissue growth factor coordinates chondrogenesis and angiogenesis during skeletal development', *Development*, vol. 130, no. 12, pp. 2779-91.

Jedsadayanmata, A., Chen, C. C., Kireeva, M. L., Lau, L. F. & Lam, S. C. 1999, 'Activation-dependent adhesion of human platelets to Cyr61 and Fisp12/mouse connective tissue growth factor is mediated through integrin alpha(IIb)beta(3)', *J Biol Chem*, vol. 274, no. 34, pp. 24321-7.

Jeha, S., Pei, D., Raimondi, S. C., Onciu, M., Campana, D., Cheng, C., Sandlund, J. T., Ribeiro, R. C., Rubnitz, J. E., Howard, S. C., Downing, J. R., Evans, W. E., Relling, M. V. & Pui, C. H. 2009, 'Increased risk for CNS relapse in pre-B cell leukemia with the t(1;19)/TCF3-PBX1', *Leukemia*, vol. 23, no. 8, pp. 1406-9.

Jude, C. D., Climer, L., Xu, D., Artinger, E., Fisher, J. K. & Ernst, P. 2007, 'Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors', *Cell Stem Cell*, vol. 1, no. 3, pp. 324-37.

Kaatsch, P. 2010, 'Epidemiology of childhood cancer', *Cancer Treat Rev*, vol. 36, no. 4, pp. 277-85.

Kalluri, R. & Zeisberg, M. 2006, 'Fibroblasts in cancer', *Nat Rev Cancer*, vol. 6, no. 5, pp. 392-401.

Kang, H., Chen, I. M., Wilson, C. S., Bedrick, E. J., Harvey, R. C., Atlas, S. R., Devidas, M., Mullighan, C. G., Wang, X., Murphy, M., Ar, K., Wharton, W.,

Borowitz, M. J., Bowman, W. P., Bhojwani, D., Carroll, W. L., Camitta, B. M., Reaman, G. H., Smith, M. A., Downing, J. R., Hunger, S. P. & Willman, C. L. 2010, 'Gene expression classifiers for relapse-free survival and minimal residual disease improve risk classification and outcome prediction in pediatric B-precursor acute lymphoblastic leukemia', *Blood*, vol. 115, no. 7, pp. 1394-405.

Kang, Y., Siegel, P. M., Shu, W., Drobnjak, M., Kakonen, S. M., Cordon-Cardo, C., Guise, T. A. & Massague, J. 2003, 'A multigenic program mediating breast cancer metastasis to bone', *Cancer Cell*, vol. 3, no. 6, pp. 537-49.

Kawaguchi, Y., Ota, Y., Kawamoto, M., Ito, I., Tsuchiya, N., Sugiura, T., Katsumata, Y., Soejima, M., Sato, S., Hasegawa, M., Fujimoto, M., Takehara, K., Kuwana, M., Yamanaka, H. & Hara, M. 2009, 'Association study of a polymorphism of the CTGF gene and susceptibility to systemic sclerosis in the Japanese population', *Ann Rheum Dis*, vol. 68, no. 12, pp. 1921-4.

Kees, U. R. 1987, 'Resistance to 1-beta-D-arabinofuranosylcytosine after high-dose treatment childhood lymphoblastic leukemia: isolation of a drug resistant and a sensitive cell line', *Cancer Res*, vol. 47, no. 12, pp. 3088-91.

Kees, U. R., Campbell, L. J., Ford, J., Willoughby, M. L., Peroni, S. E., Ranford, P. R. & Garson, O. M. 1995, 'New translocation t(2;13)(p12;q34) and rearrangement of the MLL gene in a childhood leukemia cell line', *Genes Chromosomes Cancer*, vol. 12, no. 3, pp. 201-8.

Kees, U. R., Ford, J., Price, P. J., Meyer, B. F. & Herrmann, R. P. 1987, 'PER-117: a new human ALL cell line with an immature thymic phenotype', *Leuk Res*, vol. 11, no. 5, pp. 489-98.

Kees, U. R., Ford, J., Watson, M., Murch, A., Ringner, M., Walker, R. L. & Meltzer, P. 2003, 'Gene expression profiles in a panel of childhood leukemia cell lines mirror critical features of the disease', *Mol Cancer Ther*, vol. 2, no. 7, pp. 671-7.

Kees, U. R., Lukeis, R., Ford, J., Willoughby, M. L. & Garson, O. M. 1990, 'Establishment and characterization of a childhood pre-B acute lymphoblastic leukemia cell line, PER-278, with chromosome translocations t(1;19) and t(1;9)', *Cancer Genet Cytogenet*, vol. 46, no. 2, pp. 201-8.

Kikuchi, R., Tsuda, H., Kanai, Y., Kasamatsu, T., Sengoku, K., Hirohashi, S., Inazawa, J. & Imoto, I. 2007, 'Promoter hypermethylation contributes to frequent inactivation of a putative conditional tumor suppressor gene connective tissue growth factor in ovarian cancer', *Cancer Res*, vol. 67, no. 15, pp. 7095-105.

Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. 1997, 'Identification of a family of low-affinity insulin-like growth factor binding proteins (IGFBPs): characterization of connective tissue growth factor as a

member of the IGFBP superfamily', *Proc Natl Acad Sci U S A*, vol. 94, no. 24, pp. 12981-6.

Kim, J., Sif, S., Jones, B., Jackson, A., Koipally, J., Heller, E., Winandy, S., Viel, A., Sawyer, A., Ikeda, T., Kingston, R. & Georgopoulos, K. 1999, 'Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes', *Immunity*, vol. 10, no. 3, pp. 345-55.

Kim, M., Yim, S. H., Cho, N. S., Kang, S. H., Ko, D. H., Oh, B., Kim, T. Y., Min, H. J., She, C. J., Kang, H. J., Shin, H. Y., Ahn, H. S., Yoon, S. S., Kim, B. K., Shin, H. R., Han, K. S., Cho, H. I. & Lee, D. S. 2009, 'Homozygous deletion of CDKN2A (p16, p14) and CDKN2B (p15) genes is a poor prognostic factor in adult but not in childhood B-lineage acute lymphoblastic leukemia: a comparative deletion and hypermethylation study', *Cancer Genet Cytogenet*, vol. 195, no. 1, pp. 59-65.

Koipally, J., Kim, J., Jones, B., Jackson, A., Avitahl, N., Winandy, S., Trevisan, M., Nichogiannopoulou, A., Kelley, C. & Georgopoulos, K. 1999, 'Ikaros chromatin remodeling complexes in the control of differentiation of the hemo-lymphoid system', *Cold Spring Harb Symp Quant Biol*, vol. 64, pp. 79-86.

Koliopanos, A., Friess, H., di Mola, F. F., Tang, W. H., Kubulus, D., Brigstock, D., Zimmermann, A. & Buchler, M. W. 2002, 'Connective tissue growth factor gene expression alters tumor progression in esophageal cancer', *World J Surg*, vol. 26, no. 4, pp. 420-7.

Komorowsky, C., Ocker, M. & Goppelt-Struebe, M. 2009, 'Differential regulation of connective tissue growth factor in renal cells by histone deacetylase inhibitors', *J Cell Mol Med*, vol. 13, no. 8B, pp. 2353-64.

Kondo, S., Kubota, S., Mukudai, Y., Moritani, N., Nishida, T., Matsushita, H., Matsumoto, S., Sugahara, T. & Takigawa, M. 2006, 'Hypoxic regulation of stability of connective tissue growth factor/CCN2 mRNA by 3'-untranslated region interacting with a cellular protein in human chondrosarcoma cells', *Oncogene*, vol. 25, no. 7, pp. 1099-110.

Kondo, Y. & Issa, J. P. 2010, 'DNA methylation profiling in cancer', *Expert Rev Mol Med*, vol. 12, p. e23.

Kouzarides, T. 2007, 'Chromatin modifications and their function', *Cell*, vol. 128, no. 4, pp. 693-705.

Kriaucionis, S. & Heintz, N. 2009, 'The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain', *Science*, vol. 324, no. 5929, pp. 929-30.

Kroening, S., Neubauer, E., Wessel, J., Wiesener, M. & Goppelt-Struebe, M. 2009, 'Hypoxia interferes with connective tissue growth factor (CTGF) gene expression in

human proximal tubular cell lines', *Nephrol Dial Transplant*, vol. 24, no. 11, pp. 3319-25.

Krug, U., Ganser, A. & Koeffler, H. P. 2002, 'Tumor suppressor genes in normal and malignant hematopoiesis', *Oncogene*, vol. 21, no. 21, pp. 3475-95.

Kubo, M., Kikuchi, K., Nashiro, K., Kakinuma, T., Hayashi, N., Nanko, H. & Tamaki, K. 1998, 'Expression of fibrogenic cytokines in desmoplastic malignant melanoma', *Br J Dermatol*, vol. 139, no. 2, pp. 192-7.

Kubota, S., Kondo, S., Eguchi, T., Hattori, T., Nakanishi, T., Pomerantz, R. J. & Takigawa, M. 2000, 'Identification of an RNA element that confers post-transcriptional repression of connective tissue growth factor/hypertrophic chondrocyte specific 24 (ctgf/hcs24) gene: similarities to retroviral RNA-protein interactions', *Oncogene*, vol. 19, no. 41, pp. 4773-86.

Kubota, S., Mukudai, Y., Moritani, N. H., Nakao, K., Kawata, K. & Takigawa, M. 2005, 'Translational repression by the cis-acting element of structure-anchored repression (CAESAR) of human ctgf/ccn2 mRNA', *FEBS Lett*, vol. 579, no. 17, pp. 3751-8.

Kuiper, E. J., Hughes, J. M., Van Geest, R. J., Vogels, I. M., Goldschmeding, R., Van Noorden, C. J., Schlingemann, R. O. & Klaassen, I. 2007a, 'Effect of VEGF-A on expression of profibrotic growth factor and extracellular matrix genes in the retina', *Invest Ophthalmol Vis Sci*, vol. 48, no. 9, pp. 4267-76.

Kuiper, R. P., Schoenmakers, E. F., van Reijmersdal, S. V., Hehir-Kwa, J. Y., van Kessel, A. G., van Leeuwen, F. N. & Hoogerbrugge, P. M. 2007b, 'High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression', *Leukemia*, vol. 21, no. 6, pp. 1258-66.

Kuiper, R. P., Waanders, E., van der Velden, V. H., van Reijmersdal, S. V., Venkatachalam, R., Scheijen, B., Sonneveld, E., van Dongen, J. J., Veerman, A. J., van Leeuwen, F. N., van Kessel, A. G. & Hoogerbrugge, P. M. 2010, 'IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL', *Leukemia*, vol. 24, no. 7, pp. 1258-64.

Kundel, D. W., Brecher, G., Bodey, G. P. & Brittin, G. M. 1964, 'Reticulin Fibrosis and Bone Infarction in Acute Leukemia. Implications for Prognosis', *Blood*, vol. 23, pp. 526-44.

Lane, S. W., Scadden, D. T. & Gilliland, D. G. 2009, 'The leukemic stem cell niche: current concepts and therapeutic opportunities', *Blood*, vol. 114, no. 6, pp. 1150-7.

Lang, C., Sauter, M., Szalay, G., Racchi, G., Grassi, G., Rainaldi, G., Mercatanti, A., Lang, F., Kandolf, R. & Klingel, K. 2007, 'Connective tissue growth factor: a

crucial cytokine-mediating cardiac fibrosis in ongoing enterovirus myocarditis', *J Mol Med*.

Laurila, K. & Lahdesmaki, H. 2009, 'Systematic analysis of disease-related regulatory mutation classes reveals distinct effects on transcription factor binding', *In Silico Biol*, vol. 9, no. 4, pp. 209-24.

Leask, A. 2004, 'Transcriptional profiling of the scleroderma fibroblast reveals a potential role for connective tissue growth factor (CTGF) in pathological fibrosis', *Keio J Med*, vol. 53, no. 2, pp. 74-7.

Leask, A. 2010, 'Potential therapeutic targets for cardiac fibrosis: TGFbeta, angiotensin, endothelin, CCN2, and PDGF, partners in fibroblast activation', *Circ Res*, vol. 106, no. 11, pp. 1675-80.

Leask, A. & Abraham, D. J. 2006, 'All in the CCN family: essential matricellular signaling modulators emerge from the bunker', *J Cell Sci*, vol. 119, no. Pt 23, pp. 4803-10.

Lee, C. H., Shah, B., Muioli, E. K. & Mao, J. J. 2010, 'CTGF directs fibroblast differentiation from human mesenchymal stem/stromal cells and defines connective tissue healing in a rodent injury model', *J Clin Invest*, vol. 120, no. 9, pp. 3340-9.

Lee, J. H. & Skalnik, D. G. 2005, 'CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex', *J Biol Chem*, vol. 280, no. 50, pp. 41725-31.

Li, B., Jackson, J., Simon, M. D., Fleharty, B., Gogol, M., Seidel, C., Workman, J. L. & Shilatifard, A. 2009a, 'Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription', *J Biol Chem*, vol. 284, no. 12, pp. 7970-6.

Li, X., Zhao, L., Jiang, H. & Wang, W. 2009b, 'Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes', *J Mol Evol*, vol. 68, no. 1, pp. 56-65.

Li, Y., Lu, Y., Ceng, Y. & Yang, X. 2007, '[Expression of connective tissue growth factor (CTGF), osteopontin (OPN) and clinical significances in the laryngeal squamous cell carcinoma tissues]', *Lin Chung Er Bi Yan Hou Tou Jing Wai Ke Za Zhi*, vol. 21, no. 3, pp. 121-3.

Lin, B. R., Chang, C. C., Che, T. F., Chen, S. T., Chen, R. J., Yang, C. Y., Jeng, Y. M., Liang, J. T., Lee, P. H., Chang, K. J., Chau, Y. P. & Kuo, M. L. 2005, 'Connective tissue growth factor inhibits metastasis and acts as an independent prognostic marker in colorectal cancer', *Gastroenterology*, vol. 128, no. 1, pp. 9-23.

Lind, G. E., Thorstensen, L., Lovig, T., Meling, G. I., Hamelin, R., Rognum, T. O., Esteller, M. & Lothe, R. A. 2004, 'A CpG island hypermethylation profile of primary colorectal carcinomas and colon cancer cell lines', *Mol Cancer*, vol. 3, p. 28.

Lionetti, M., Biasiolo, M., Agnelli, L., Todoerti, K., Mosca, L., Fabris, S., Sales, G., Deliliers, G. L., Biciato, S., Lombardi, L., Bortoluzzi, S. & Neri, A. 2009, 'Identification of microRNA expression patterns and definition of a microRNA/mRNA regulatory network in distinct molecular groups of multiple myeloma', *Blood*, vol. 114, no. 25, pp. e20-6.

Lo Celso, C., Fleming, H. E., Wu, J. W., Zhao, C. X., Miake-Lye, S., Fujisaki, J., Cote, D., Rowe, D. W., Lin, C. P. & Scadden, D. T. 2009, 'Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche', *Nature*, vol. 457, no. 7225, pp. 92-6.

Luscher, B. & Larsson, L. G. 1999, 'The basic region/helix-loop-helix/leucine zipper domain of Myc proto-oncoproteins: function and regulation', *Oncogene*, vol. 18, no. 19, pp. 2955-66.

Maia, A. T., van der Velden, V. H., Harrison, C. J., Szczepanski, T., Williams, M. D., Griffiths, M. J., van Dongen, J. J. & Greaves, M. F. 2003, 'Prenatal origin of hyperdiploid acute lymphoblastic leukemia in identical twins', *Leukemia*, vol. 17, no. 11, pp. 2202-6.

Manetti, M., Neumann, E., Milia, A. F., Tarner, I. H., Bechi, P., Matucci-Cerinic, M., Ibba-Manneschi, L. & Muller-Ladner, U. 2007, 'Severe fibrosis and increased expression of fibrogenic cytokines in the gastric wall of systemic sclerosis patients', *Arthritis Rheum*, vol. 56, no. 10, pp. 3442-3447.

Mansson, R., Tsapogas, P., Akerlund, M., Lagergren, A., Gisler, R. & Sigvardsson, M. 2004, 'Pearson correlation analysis of microarray data allows for the identification of genetic targets for early B-cell factor', *J Biol Chem*, vol. 279, no. 17, pp. 17905-13.

Mansson, R., Zandi, S., Bryder, D. & Sigvardsson, M. 2009, 'The Road to Commitment: Lineage Restriction Events in Hematopoiesis', in *Molecular Basis of Hematopoiesis*, eds A. Wickrema & B. Kee, Springer, New York, pp. 23-46.

Markl, I. D., Cheng, J., Liang, G., Shibata, D., Laird, P. W. & Jones, P. A. 2001, 'Global and gene-specific epigenetic patterns in human bladder cancer genomes are relatively stable in vivo and in vitro over time', *Cancer Res*, vol. 61, no. 15, pp. 5875-84.

Marks, D. I., Kurz, B. W., Link, M. P., Ng, E., Shuster, J. J., Lauer, S. J., Carroll, D., Brodsky, I. & Haines, D. S. 1997, 'Altered expression of p53 and mdm-2 proteins at diagnosis is associated with early treatment failure in childhood acute lymphoblastic leukemia', *J Clin Oncol*, vol. 15, no. 3, pp. 1158-62.

Martin-Subero, J. I., Ammerpohl, O., Bibikova, M., Wickham-Garcia, E., Agirre, X., Alvarez, S., Bruggemann, M., Bug, S., Calasanz, M. J., Deckert, M., Dreyling, M., Du, M. Q., Durig, J., Dyer, M. J., Fan, J. B., Gesk, S., Hansmann, M. L., Harder, L., Hartmann, S., Klapper, W., Kuppers, R., Montesinos-Rongen, M., Nagel, I., Pott, C., Richter, J., Roman-Gomez, J., Seifert, M., Stein, H., Suela, J., Trumper, L., Vater, I., Prosper, F., Haferlach, C., Cruz Cigudosa, J. & Siebert, R. 2009, 'A comprehensive microarray-based DNA methylation study of 367 hematological neoplasms', *PLoS One*, vol. 4, no. 9, p. e6986.

Martinerie, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. 1992, 'Physical mapping of human loci homologous to the chicken *nov* proto-oncogene', *Oncogene*, vol. 7, no. 12, pp. 2529-34.

Mayr, C. & Bartel, D. P. 2009, 'Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells', *Cell*, vol. 138, no. 4, pp. 673-84.

McNally, R. J. & Parker, L. 2006, 'Environmental factors and childhood acute leukemias and lymphomas', *Leuk Lymphoma*, vol. 47, no. 4, pp. 583-98.

McNeer, J. L. & Nachman, J. B. 2010, 'The optimal use of steroids in paediatric acute lymphoblastic leukaemia: no easy answers', *Br J Haematol*, vol. 149, no. 5, pp. 638-52.

McNeil, D. E., Cote, T. R., Clegg, L. & Mauer, A. 2002, 'SEER update of incidence and trends in pediatric malignancies: acute lymphoblastic leukemia', *Med Pediatr Oncol*, vol. 39, no. 6, pp. 554-7; discussion 552-3.

Meads, M. B., Hazlehurst, L. A. & Dalton, W. S. 2008, 'The bone marrow microenvironment as a tumor sanctuary and contributor to drug resistance', *Clin Cancer Res*, vol. 14, no. 9, pp. 2519-26.

Meleshko, A. N., Movchan, L. V., Belevtsev, M. V. & Savitskaja, T. V. 2008, 'Relative expression of different Ikaros isoforms in childhood acute leukemia', *Blood Cells Mol Dis*, vol. 41, no. 3, pp. 278-83.

Melrose, J., Hayes, A. J., Whitelock, J. M. & Little, C. B. 2008, 'Perlecan, the "jack of all trades" proteoglycan of cartilaginous weight-bearing connective tissues', *Bioessays*, vol. 30, no. 5, pp. 457-69.

Mendez-Ferrer, S., Michurina, T. V., Ferraro, F., Mazloom, A. R., Macarthur, B. D., Lira, S. A., Scadden, D. T., Ma'ayan, A., Enikolopov, G. N. & Frenette, P. S. 2010, 'Mesenchymal and haematopoietic stem cells form a unique bone marrow niche', *Nature*, vol. 466, no. 7308, pp. 829-34.

Mercurio, S., Latinkic, B., Itasaki, N., Krumlauf, R. & Smith, J. C. 2004, 'Connective-tissue growth factor modulates WNT signalling and interacts with the WNT receptor complex', *Development*, vol. 131, no. 9, pp. 2137-47.

Messeguer, X., Escudero, R., Farre, D., Nunez, O., Martinez, J. & Alba, M. M. 2002, 'PROMO: detection of known transcription regulatory elements using species-tailored searches', *Bioinformatics*, vol. 18, no. 2, pp. 333-4.

Meyer, C., Kowarz, E., Hofmann, J., Renneville, A., Zuna, J., Trka, J., Ben Abdelali, R., Macintyre, E., De Braekeleer, E., De Braekeleer, M., Delabesse, E., de Oliveira, M. P., Cave, H., Clappier, E., van Dongen, J. J., Balgobind, B. V., van den Heuvel-Eibrink, M. M., Beverloo, H. B., Panzer-Grumayer, R., Teigler-Schlegel, A., Harbott, J., Kjeldsen, E., Schnittger, S., Koehl, U., Gruhn, B., Heidenreich, O., Chan, L. C., Yip, S. F., Krzywinski, M., Eckert, C., Moricke, A., Schrappe, M., Alonso, C. N., Schafer, B. W., Krauter, J., Lee, D. A., Zur Stadt, U., Te Kronnie, G., Sutton, R., Izraeli, S., Trakhtenbrot, L., Lo Nigro, L., Tsauro, G., Fechina, L., Szczepanski, T., Strehl, S., Ilencikova, D., Molkentin, M., Burmeister, T., Dinger, T., Klingebiel, T. & Marschalek, R. 2009, 'New insights to the MLL recombinome of acute leukemias', *Leukemia*, vol. 23, no. 8, pp. 1490-9.

Milani, L., Lundmark, A., Kiialainen, A., Nordlund, J., Flaegstad, T., Forestier, E., Heyman, M., Jonmundsson, G., Kanerva, J., Schmiegelow, K., Soderhall, S., Gustafsson, M. G., Lonnerholm, G. & Syvanen, A. C. 2010, 'DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia', *Blood*, vol. 115, no. 6, pp. 1214-25.

Mitelman, F., Johansson, B. & Mertens, F. 2010, *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (2010)*. Accessed: Jan 2010

Molnar, A. & Georgopoulos, K. 1994, 'The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins', *Mol Cell Biol*, vol. 14, no. 12, pp. 8292-303.

Moritani, N. H., Kubota, S., Nishida, T., Kawaki, H., Kondo, S., Sugahara, T. & Takigawa, M. 2003, 'Suppressive effect of overexpressed connective tissue growth factor on tumor cell growth in a human oral squamous cell carcinoma-derived cell line', *Cancer Lett*, vol. 192, no. 2, pp. 205-14.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. 2008, 'Mapping and quantifying mammalian transcriptomes by RNA-Seq', *Nat Methods*, vol. 5, no. 7, pp. 621-8.

Muehlich, S., Cicha, I., Garlich, C. D., Krueger, B., Posern, G. & Goppelt-Strube, M. 2007, 'Actin-dependent regulation of connective tissue growth factor', *Am J Physiol Cell Physiol*, vol. 292, no. 5, pp. C1732-8.

Mullighan, C. G. & Downing, J. R. 2009, 'Genome-wide profiling of genetic alterations in acute lymphoblastic leukemia: recent insights and future directions', *Leukemia*, vol. 23, no. 7, pp. 1209-18.

Mullighan, C. G., Goorha, S., Radtke, I., Miller, C. B., Coustan-Smith, E., Dalton, J. D., Girtman, K., Mathew, S., Ma, J., Pounds, S. B., Su, X., Pui, C. H., Relling, M.

V., Evans, W. E., Shurtleff, S. A. & Downing, J. R. 2007, 'Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia', *Nature*, vol. 446, no. 7137, pp. 758-64.

Mullighan, C. G., Miller, C. B., Radtke, I., Phillips, L. A., Dalton, J., Ma, J., White, D., Hughes, T. P., Le Beau, M. M., Pui, C. H., Relling, M. V., Shurtleff, S. A. & Downing, J. R. 2008, 'BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros', *Nature*, vol. 453, no. 7191, pp. 110-4.

Mullighan, C. G., Su, X., Zhang, J., Radtke, I., Phillips, L. A., Miller, C. B., Ma, J., Liu, W., Cheng, C., Schulman, B. A., Harvey, R. C., Chen, I. M., Clifford, R. J., Carroll, W. L., Reaman, G., Bowman, W. P., Devidas, M., Gerhard, D. S., Yang, W., Relling, M. V., Shurtleff, S. A., Campana, D., Borowitz, M. J., Pui, C. H., Smith, M., Hunger, S. P., Willman, C. L. & Downing, J. R. 2009, 'Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia', *N Engl J Med*, vol. 360, no. 5, pp. 470-80.

Murphy, G. & Nagase, H. 2008, 'Progress in matrix metalloproteinase research', *Mol Aspects Med*, vol. 29, no. 5, pp. 290-308.

Nagasawa, T. 2006, 'Microenvironmental niches in the bone marrow required for B-cell development', *Nat Rev Immunol*, vol. 6, no. 2, pp. 107-16.

Nagatoshi, Y., Matsuzaki, A., Suminoe, A., Inada, H., Ueda, K., Kawakami, K., Yanai, F., Nakayama, H., Moritake, H., Itonaga, N., Hotta, N., Fujita, K., Hidaka, Y., Yamanaka, T., Kawano, Y. & Okamura, J. 2010, 'Randomized trial to compare LSA2L2-type maintenance therapy to daily 6-mercaptopurine and weekly methotrexate with vincristine and dexamethasone pulse for children with acute lymphoblastic leukemia', *Pediatr Blood Cancer*, vol. 55, no. 2, pp. 239-47.

Ness, K. K., Morris, E. B., Nolan, V. G., Howell, C. R., Gilchrist, L. S., Stovall, M., Cox, C. L., Klosky, J. L., Gajjar, A. & Neglia, J. P. 2010, 'Physical performance limitations among adult survivors of childhood brain tumors', *Cancer*, vol. 116, no. 12, pp. 3034-44.

Nestor, C., Ruzov, A., Meehan, R. & Dunican, D. 2010, 'Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA', *Biotechniques*, vol. 48, no. 4, pp. 317-9.

Nguyen, T. Q., Roestenberg, P., van Nieuwenhoven, F. A., Bovenschen, N., Li, Z., Xu, L., Oliver, N., Aten, J., Joles, J. A., Vial, C., Brandan, E., Lyons, K. M. & Goldschmeding, R. 2008, 'CTGF inhibits BMP-7 signaling in diabetic nephropathy', *J Am Soc Nephrol*, vol. 19, no. 11, pp. 2098-107.

Nie, L., Wu, G. & Zhang, W. 2006, 'Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in *Desulfovibrio vulgaris*: a quantitative analysis', *Genetics*, vol. 174, no. 4, pp. 2229-43.

- Nielsen, J. S. & McNagny, K. M. 2009, 'CD34 is a key regulator of hematopoietic stem cell trafficking to bone marrow and mast cell progenitor trafficking in the periphery', *Microcirculation*, vol. 16, no. 6, pp. 487-96.
- Niitsu, N., Hayashi, Y., Sugita, K. & Honma, Y. 2001, 'Sensitization by 5-aza-2'-deoxycytidine of leukaemia cells with MLL abnormalities to induction of differentiation by all-trans retinoic acid and 1alpha,25-dihydroxyvitamin D3', *Br J Haematol*, vol. 112, no. 2, pp. 315-26.
- Nishida, T., Kawaki, H., Baxter, R. M., Deyoung, R. A., Takigawa, M. & Lyons, K. M. 2007, 'CCN2 (Connective Tissue Growth Factor) is essential for extracellular matrix production and integrin signaling in chondrocytes', *J Cell Commun Signal*, vol. 1, no. 1, pp. 45-58.
- Nishida, T., Maeda, A., Kubota, S. & Takigawa, M. 2008, 'Role of mechanical-stress inducible protein Hcs24/CTGF/CCN2 in cartilage growth and regeneration: mechanical stress induces expression of Hcs24/CTGF/CCN2 in a human chondrocytic cell line HCS-2/8, rabbit costal chondrocytes and meniscus tissue cells', *Biorheology*, vol. 45, no. 3-4, pp. 289-99.
- Nishida, T., Nakanishi, T., Asano, M., Shimo, T. & Takigawa, M. 2000, 'Effects of CTGF/Hcs24, a hypertrophic chondrocyte-specific gene product, on the proliferation and differentiation of osteoblastic cells in vitro', *J Cell Physiol*, vol. 184, no. 2, pp. 197-206.
- Noren-Nystrom, U., Roos, G., Bergh, A., Botling, J., Lonnerholm, G., Porwit, A., Heyman, M. & Forestier, E. 2008, 'Bone marrow fibrosis in childhood acute lymphoblastic leukemia correlates to biological factors, treatment response and outcome', *Leukemia*, vol. 22, no. 3, pp. 504-10.
- Oeffinger, K. C., Nathan, P. C. & Kremer, L. C. 2008, 'Challenges after curative treatment for childhood cancer and long-term follow up of survivors', *Pediatr Clin North Am*, vol. 55, no. 1, pp. 251-73, xiii.
- Ohgawara, T., Kubota, S., Kawaki, H., Kondo, S., Eguchi, T., Kurio, N., Aoyama, E., Sasaki, A. & Takigawa, M. 2009, 'Regulation of chondrocytic phenotype by micro RNA 18a: involvement of Ccn2/Ctgf as a major target gene', *FEBS Lett*, vol. 583, no. 6, pp. 1006-10.
- Okitsu, C. Y. & Hsieh, C. L. 2007, 'DNA methylation dictates histone H3K4 methylation', *Mol Cell Biol*, vol. 27, no. 7, pp. 2746-57.
- Okumura, A. J., Peterson, L. F., Lo, M. C. & Zhang, D. E. 2007, 'Expression of AML/Runx and ETO/MTG family members during hematopoietic differentiation of embryonic stem cells', *Exp Hematol*, vol. 35, no. 6, pp. 978-88.

Omatsu, Y., Sugiyama, T., Kohara, H., Kondoh, G., Fujii, N., Kohno, K. & Nagasawa, T. 2010, 'The essential functions of adipo-osteogenic progenitors as the hematopoietic stem and progenitor cell niche', *Immunity*, vol. 33, no. 3, pp. 387-99.

Onciu, M. 2009, 'Acute lymphoblastic leukemia', *Hematol Oncol Clin North Am*, vol. 23, no. 4, pp. 655-74.

Park, J., Kunjibettu, S., McMahon, S. B. & Cole, M. D. 2001, 'The ATM-related domain of TRRAP is required for histone acetyltransferase recruitment and Myc-dependent oncogenesis', *Genes Dev*, vol. 15, no. 13, pp. 1619-24.

Parrish, Y. K., Baez, I., Milford, T. A., Benitez, A., Galloway, N., Rogerio, J. W., Sahakian, E., Kagoda, M., Huang, G., Hao, Q. L., Sevilla, Y., Barsky, L. W., Zielinska, E., Price, M. A., Wall, N. R., Dovat, S. & Payne, K. J. 2009, 'IL-7 Dependence in human B lymphopoiesis increases during progression of ontogeny from cord blood to bone marrow', *J Immunol*, vol. 182, no. 7, pp. 4255-66.

Paulsson, K. & Johansson, B. 2009, 'High hyperdiploid childhood acute lymphoblastic leukemia', *Genes Chromosomes Cancer*, vol. 48, no. 8, pp. 637-60.

Peltz, S. W., He, F., Welch, E. & Jacobson, A. 1994, 'Nonsense-mediated mRNA decay in yeast', *Prog Nucleic Acid Res Mol Biol*, vol. 47, pp. 271-98.

Perbal, B. 2001, 'NOV (nephroblastoma overexpressed) and the CCN family of genes: structural and functional issues', *Mol Pathol*, vol. 54, no. 2, pp. 57-79.

Perbal, B. 2009, 'Alternative splicing of CCN mRNAs it has been upon us', *J Cell Commun Signal*, vol. 3, no. 2, pp. 153-7.

Perdomo, J., Holmes, M., Chong, B. & Crossley, M. 2000, 'Eos and pegasus, two members of the Ikaros family of proteins with distinct DNA binding activities', *J Biol Chem*, vol. 275, no. 49, pp. 38347-54.

Phanish, M. K., Wahab, N. A., Colville-Nash, P., Hendry, B. M. & Dockrell, M. E. 2006, 'The differential role of Smad2 and Smad3 in the regulation of pro-fibrotic TGFbeta1 responses in human proximal-tubule epithelial cells', *Biochem J*, vol. 393, no. Pt 2, pp. 601-7.

Pickles, M. & Leask, A. 2007, 'Analysis of CCN2 promoter activity in PANC-1 cells: regulation by ras/MEK/ERK', *J Cell Commun Signal*, vol. 1, no. 2, pp. 85-90.

Pieters, R. & Carroll, W. L. 2010, 'Biology and treatment of acute lymphoblastic leukemia', *Hematol Oncol Clin North Am*, vol. 24, no. 1, pp. 1-18.

Pomerantz, J., Schreiber-Agus, N., Liegeois, N. J., Silverman, A., Alland, L., Chin, L., Potes, J., Chen, K., Orlov, I., Lee, H. W., Cordon-Cardo, C. & DePinho, R. A. 1998, 'The Ink4a tumor suppressor gene product, p19Arf, interacts with MDM2 and neutralizes MDM2's inhibition of p53', *Cell*, vol. 92, no. 6, pp. 713-23.

Poppe, B., De Paepe, A. & Speleman, F. 2007, 'Acquired chromosomal rearrangements targeting selected transcription factors: contribution of molecular cytogenetic and expression analyses to the identification of clinically and biologically relevant subgroups in hematological malignancies', *Verh K Acad Geneeskd Belg*, vol. 69, no. 1, pp. 47-64.

Portales-Casamar, E., Thongjuea, S., Kwon, A. T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W. W. & Sandelin, A. 2010, 'JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles', *Nucleic Acids Res*, vol. 38, no. Database issue, pp. D105-10.

Pui, C. H. 2006, 'Childhood Leukemias', in, 2nd edn, Cambridge University Press, Cambridge, pp. 166-169.

Pui, C. H., Carroll, W. L., Meshinchi, S. & Arceci, R. J. 2011, 'Biology, risk stratification, and therapy of pediatric acute leukemias: an update', *J Clin Oncol*, vol. 29, no. 5, pp. 551-65.

Pui, C. H. & Evans, W. E. 2006, 'Treatment of acute lymphoblastic leukemia', *N Engl J Med*, vol. 354, no. 2, pp. 166-78.

Pui, C. H., Pei, D., Sandlund, J. T., Ribeiro, R. C., Rubnitz, J. E., Raimondi, S. C., Onciu, M., Campana, D., Kun, L. E., Jeha, S., Cheng, C., Howard, S. C., Metzger, M. L., Bhojwani, D., Downing, J. R., Evans, W. E. & Relling, M. V. 2010, 'Long-term results of St Jude Total Therapy Studies 11, 12, 13A, 13B, and 14 for childhood acute lymphoblastic leukemia', *Leukemia*, vol. 24, no. 2, pp. 371-82.

Pui, C. H., Sandlund, J. T., Pei, D., Rivera, G. K., Howard, S. C., Ribeiro, R. C., Rubnitz, J. E., Razzouk, B. I., Hudson, M. M., Cheng, C., Raimondi, S. C., Behm, F. G., Downing, J. R., Relling, M. V. & Evans, W. E. 2003, 'Results of therapy for acute lymphoblastic leukemia in black and white children', *JAMA*, vol. 290, no. 15, pp. 2001-7.

Purton, L. E. & Scadden, D. T. 2007, 'Limiting factors in murine hematopoietic stem cell assays', *Cell Stem Cell*, vol. 1, no. 3, pp. 263-70.

Raimondi, S. C., Behm, F. G., Roberson, P. K., Williams, D. L., Pui, C. H., Crist, W. M., Look, A. T. & Rivera, G. K. 1990, 'Cytogenetics of pre-B-cell acute lymphoblastic leukemia with emphasis on prognostic implications of the t(1;19)', *J Clin Oncol*, vol. 8, no. 8, pp. 1380-8.

Rebollo, A. & Schmitt, C. 2003, 'Ikaros, Aiolos and Helios: transcription regulators and lymphoid malignancies', *Immunol Cell Biol*, vol. 81, no. 3, pp. 171-5.

Recchia, A. G., Filice, E., Pellegrino, D., Dobrina, A., Cerra, M. C. & Maggiolini, M. 2009, 'Endothelin-1 induces connective tissue growth factor expression in cardiomyocytes', *J Mol Cell Cardiol*, vol. 46, no. 3, pp. 352-9.

Riggs, A. D. 1975, 'X inactivation, differentiation, and DNA methylation', *Cytogenet Cell Genet*, vol. 14, no. 1, pp. 9-25.

Rodriguez-Pinto, D. 2005, 'B cells as antigen presenting cells', *Cell Immunol*, vol. 238, no. 2, pp. 67-75.

Ross, M. E., Zhou, X., Song, G., Shurtleff, S. A., Girtman, K., Williams, W. K., Liu, H. C., Mahfouz, R., Raimondi, S. C., Lenny, N., Patel, A. & Downing, J. R. 2003, 'Classification of pediatric acute lymphoblastic leukemia by gene expression profiling', *Blood*, vol. 102, no. 8, pp. 2951-9.

Rowe, J. M., Buck, G., Burnett, A. K., Chopra, R., Wiernik, P. H., Richards, S. M., Lazarus, H. M., Franklin, I. M., Litzow, M. R., Ciobanu, N., Prentice, H. G., Durrant, J., Tallman, M. S. & Goldstone, A. H. 2005, 'Induction therapy for adults with acute lymphoblastic leukemia: results of more than 1500 patients from the international ALL trial: MRC UKALL XII/ECOG E2993', *Blood*, vol. 106, no. 12, pp. 3760-7.

Ruan, K., Fang, X. & Ouyang, G. 2009, 'MicroRNAs: novel regulators in the hallmarks of human cancer', *Cancer Lett*, vol. 285, no. 2, pp. 116-26.

Rubnitz, J. E. & Crist, W. M. 1997, 'Molecular genetics of childhood cancer: implications for pathogenesis, diagnosis, and treatment', *Pediatrics*, vol. 100, no. 1, pp. 101-8.

Rueda, B., Simeon, C., Hesselstrand, R., Herrick, A., Worthington, J., Ortego-Centeno, N., Riemekasten, G., Fonollosa, V., Vonk, M. C., van den Hoogen, F. H., Sanchez-Roman, J., Aguirre-Zamorano, M. A., Garcia-Portales, R., Pros, A., Camps, M. T., Gonzalez-Gay, M. A., Gonzalez-Escribano, M. F., Coenen, M. J., Lambert, N., Nelson, J. L., Radstake, T. R. & Martin, J. 2009, 'A large multicentre analysis of CTGF -945 promoter polymorphism does not confirm association with systemic sclerosis susceptibility or phenotype', *Ann Rheum Dis*, vol. 68, no. 10, pp. 1618-20.

Ruiz, A., Williams, O. & Brady, H. J. 2008, 'The Ikaros splice isoform, Ikaros 6, immortalizes murine haematopoietic progenitor cells', *Int J Cancer*, vol. 123, no. 6, pp. 1240-5.

Sala-Torra, O., Gundacker, H. M., Stirewalt, D. L., Ladne, P. A., Pogossova-Agadjanya, E. L., Slovak, M. L., Willman, C. L., Heimfeld, S., Boldt, D. H. & Radich, J. P. 2007, 'Connective tissue growth factor (CTGF) expression and outcome in adult patients with acute lymphoblastic leukemia', *Blood*, vol. 109, no. 7, pp. 3080-3.

Samarin, J., Wessel, J., Cicha, I., Kroening, S., Warnecke, C. & Goppelt-Struebe, M. 2010, 'FoxO proteins mediate hypoxic induction of connective tissue growth factor in endothelial cells', *J Biol Chem*, vol. 285, no. 7, pp. 4328-36.

Sarraf, S. A. & Stancheva, I. 2004, 'Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly', *Mol Cell*, vol. 15, no. 4, pp. 595-605.

Sato, S., Nagaoka, T., Hasegawa, M., Tamatani, T., Nakanishi, T., Takigawa, M. & Takehara, K. 2000, 'Serum levels of connective tissue growth factor are elevated in patients with systemic sclerosis: association with extent of skin sclerosis and severity of pulmonary fibrosis', *J Rheumatol*, vol. 27, no. 1, pp. 149-54.

Schober, J. M., Chen, N., Grzeszkiewicz, T. M., Jovanovic, I., Emeson, E. E., Ugarova, T. P., Ye, R. D., Lau, L. F. & Lam, S. C. 2002, 'Identification of integrin alpha(M)beta(2) as an adhesion receptor on peripheral blood monocytes for Cyr61 (CCN1) and connective tissue growth factor (CCN2): immediate-early gene products expressed in atherosclerotic lesions', *Blood*, vol. 99, no. 12, pp. 4457-65.

Schofield, K. P., Gallagher, J. T. & David, G. 1999, 'Expression of proteoglycan core proteins in human bone marrow stroma', *Biochem J*, vol. 343 Pt 3, pp. 663-8.

Segarini, P. R., Nesbitt, J. E., Li, D., Hays, L. G., Yates, J. R., 3rd & Carmichael, D. F. 2001, 'The low density lipoprotein receptor-related protein/alpha2-macroglobulin receptor is a receptor for connective tissue growth factor', *J Biol Chem*, vol. 276, no. 44, pp. 40659-67.

Seifert, H. H., Schmiemann, V., Mueller, M., Kazimirek, M., Onofre, F., Neuhausen, A., Florl, A. R., Ackermann, R., Boecking, A., Schulz, W. A. & Grote, H. J. 2007, 'In situ detection of global DNA hypomethylation in exfoliative urine cytology of patients with suspected bladder cancer', *Exp Mol Pathol*, vol. 82, no. 3, pp. 292-7.

Shakunaga, T., Ozaki, T., Ohara, N., Asaumi, K., Doi, T., Nishida, K., Kawai, A., Nakanishi, T., Takigawa, M. & Inoue, H. 2000, 'Expression of connective tissue growth factor in cartilaginous tumors', *Cancer*, vol. 89, no. 7, pp. 1466-73.

Shi-Wen, X., Leask, A. & Abraham, D. 2008, 'Regulation and function of connective tissue growth factor/CCN2 in tissue repair, scarring and fibrosis', *Cytokine Growth Factor Rev*, vol. 19, no. 2, pp. 133-44.

Shi-wen, X., Stanton, L. A., Kennedy, L., Pala, D., Chen, Y., Howat, S. L., Renzoni, E. A., Carter, D. E., Bou-Gharios, G., Stratton, R. J., Pearson, J. D., Beier, F., Lyons, K. M., Black, C. M., Abraham, D. J. & Leask, A. 2006, 'CCN2 is necessary for adhesive responses to transforming growth factor-beta1 in embryonic fibroblasts', *J Biol Chem*, vol. 281, no. 16, pp. 10715-26.

Shimo, T., Kubota, S., Yoshioka, N., Ibaragi, S., Isowa, S., Eguchi, T., Sasaki, A. & Takigawa, M. 2006, 'Pathogenic role of connective tissue growth factor (CTGF/CCN2) in osteolytic metastasis of breast cancer', *J Bone Miner Res*, vol. 21, no. 7, pp. 1045-59.

Sidhu, S., Martin, E., Gicquel, C., Melki, J., Clark, S. J., Campbell, P., Magarey, C. J., Schulte, K. M., Roher, H. D., Delbridge, L. & Robinson, B. G. 2005, 'Mutation and methylation analysis of TP53 in adrenal carcinogenesis', *Eur J Surg Oncol*, vol. 31, no. 5, pp. 549-54.

Skorski, T. 2002, 'BCR/ABL regulates response to DNA damage: the role in resistance to genotoxic treatment and in genomic instability', *Oncogene*, vol. 21, no. 56, pp. 8591-604.

Smith, E. & Shilatifard, A. 2011, 'The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes', *Mol Cell*, vol. 40, no. 5, pp. 689-701.

Smith, I. M., Glazer, C. A., Mithani, S. K., Ochs, M. F., Sun, W., Bhan, S., Vostrov, A., Abdullaev, Z., Lobanenkov, V., Gray, A., Liu, C., Chang, S. S., Ostrow, K. L., Westra, W. H., Begum, S., Dhara, M. & Califano, J. 2009, 'Coordinated activation of candidate proto-oncogenes and cancer testes antigens via promoter demethylation in head and neck cancer and lung cancer', *PLoS One*, vol. 4, no. 3, p. e4961.

Smith, M., Arthur, D., Camitta, B., Carroll, A. J., Crist, W., Gaynon, P., Gelber, R., Heerema, N., Korn, E. L., Link, M., Murphy, S., Pui, C. H., Pullen, J., Reamon, G., Sallan, S. E., Sather, H., Shuster, J., Simon, R., Trigg, M., Tubergen, D., Uckun, F. & Ungerleider, R. 1996, 'Uniform approach to risk classification and treatment assignment for children with acute lymphoblastic leukemia', *J Clin Oncol*, vol. 14, no. 1, pp. 18-24.

Snowden, A. W., Gregory, P. D., Case, C. C. & Pabo, C. O. 2002, 'Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo', *Curr Biol*, vol. 12, no. 24, pp. 2159-66.

Somervaille, T. C. & Cleary, M. L. 2010, 'Grist for the MLL: how do MLL oncogenic fusion proteins generate leukemia stem cells?' *Int J Hematol*, vol. 91, no. 5, pp. 735-41.

Song, J. J., Aswad, R., Kanaan, R. A., Rico, M. C., Owen, T. A., Barbe, M. F., Safadi, F. F. & Popoff, S. N. 2007, 'Connective tissue growth factor (CTGF) acts as a downstream mediator of TGF-beta1 to induce mesenchymal cell condensation', *J Cell Physiol*, vol. 210, no. 2, pp. 398-410.

Soreq, L., Gilboa-Geffen, A., Berrih-Aknin, S., Lacoste, P., Darvasi, A., Soreq, E., Bergman, H. & Soreq, H. 2008, 'Identifying alternative hyper-splicing signatures in MG-thymoma by exon arrays', *PLoS One*, vol. 3, no. 6, p. e2392.

Stark, B., Avrahami, G., Nirel, R., Abramov, A., Attias, D., Ballin, A., Bielorai, B., Burstein, Y., Gavriel, H., Elhasid, R., Kapelushnik, J., Sthoeger, D., Toren, A., Wientraub, M., Yaniv, I. & Izraeli, S. 2009, 'Extended triple intrathecal therapy in

children with T-cell acute lymphoblastic leukaemia: a report from the Israeli National ALL-Studies', *Br J Haematol*, vol. 147, no. 1, pp. 113-24.

Steffen, C. L., Ball-Mirth, D. K., Harding, P. A., Bhattacharyya, N., Pillai, S. & Brigstock, D. R. 1998, 'Characterization of cell-associated and soluble forms of connective tissue growth factor (CTGF) produced by fibroblast cells in vitro', *Growth Factors*, vol. 15, no. 3, pp. 199-213.

Stielow, B., Kruger, I., Diezko, R., Finkernagel, F., Gillemans, N., Kong-a-San, J., Philipsen, S. & Suske, G. 2010, 'Epigenetic silencing of spermatocyte-specific and neuronal genes by SUMO modification of the transcription factor Sp3', *PLoS Genet*, vol. 6, no. 11, p. e1001203.

Stielow, B., Sapetschnig, A., Wink, C., Kruger, I. & Suske, G. 2008, 'SUMO-modified Sp3 represses transcription by provoking local heterochromatic gene silencing', *EMBO Rep*, vol. 9, no. 9, pp. 899-906.

Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. 2006, 'Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches', *Immunity*, vol. 25, no. 6, pp. 977-88.

Sun, L., Liu, A. & Georgopoulos, K. 1996, 'Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development', *EMBO J*, vol. 15, no. 19, pp. 5358-69.

Sutcliffe, M. J., Shuster, J. J., Sather, H. N., Camitta, B. M., Pullen, J., Schultz, K. R., Borowitz, M. J., Gaynon, P. S., Carroll, A. J. & Heerema, N. A. 2005, 'High concordance from independent studies by the Children's Cancer Group (CCG) and Pediatric Oncology Group (POG) associating favorable prognosis with combined trisomies 4, 10, and 17 in children with NCI Standard-Risk B-precursor Acute Lymphoblastic Leukemia: a Children's Oncology Group (COG) initiative', *Leukemia*, vol. 19, no. 5, pp. 734-40.

Suzuma, K., Naruse, K., Suzuma, I., Takahara, N., Ueki, K., Aiello, L. P. & King, G. L. 2000, 'Vascular endothelial growth factor induces expression of connective tissue growth factor via KDR, Flt1, and phosphatidylinositol 3-kinase-akt-dependent pathways in retinal vascular cells', *J Biol Chem*, vol. 275, no. 52, pp. 40725-31.

Tabayashi, T., Ishimaru, F., Takata, M., Kataoka, I., Nakase, K., Kozuka, T. & Tanimoto, M. 2007, 'Characterization of the short isoform of Helios overexpressed in patients with T-cell malignancies', *Cancer Sci*, vol. 98, no. 2, pp. 182-8.

Takai, D. & Jones, P. A. 2003, 'The CpG island searcher: a new WWW resource', *In Silico Biol*, vol. 3, no. 3, pp. 235-40.

Takigawa, M., Nakanishi, T., Kubota, S. & Nishida, T. 2003, 'Role of CTGF/HCS24/ecogenin in skeletal growth control', *J Cell Physiol*, vol. 194, no. 3, pp. 256-66.

Tan, T. W., Lai, C. H., Huang, C. Y., Yang, W. H., Chen, H. T., Hsu, H. C., Fong, Y. C. & Tang, C. H. 2009, 'CTGF enhances migration and MMP-13 up-regulation via alphavbeta3 integrin, FAK, ERK, and NF-kappaB-dependent pathway in human chondrosarcoma cells', *J Cell Biochem*, vol. 107, no. 2, pp. 345-56.

Tanaka, M., Oikawa, K., Takanashi, M., Kudo, M., Ohyashiki, J., Ohyashiki, K. & Kuroda, M. 2009, 'Down-regulation of miR-92 in human plasma is a novel marker for acute leukemia patients', *PLoS One*, vol. 4, no. 5, p. e5532.

Tanaka, S., Sugimachi, K., Saeki, H., Kinoshita, J., Ohga, T., Shimada, M., Maehara, Y. & Sugimachi, K. 2001, 'A novel variant of WISP1 lacking a Von Willebrand type C module overexpressed in scirrhous gastric carcinoma', *Oncogene*, vol. 20, no. 39, pp. 5525-32.

Thelander, E. F., Ichimura, K., Corcoran, M., Barbany, G., Nordgren, A., Heyman, M., Berglund, M., Mungall, A., Rosenquist, R., Collins, V. P., Grandner, D., Larsson, C. & Lagercrantz, S. 2008, 'Characterization of 6q deletions in mature B cell lymphomas and childhood acute lymphoblastic leukemia', *Leuk Lymphoma*, vol. 49, no. 3, pp. 477-87.

Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., Yi, E. C., Dai, H., Thorsson, V., Eng, J., Goodlett, D., Berger, J. P., Gunter, B., Linseley, P. S., Stoughton, R. B., Aebersold, R., Collins, S. J., Hanlon, W. A. & Hood, L. E. 2004, 'Integrated genomic and proteomic analyses of gene expression in Mammalian cells', *Mol Cell Proteomics*, vol. 3, no. 10, pp. 960-9.

Tonelli, R., Sartini, R., Fronza, R., Freccero, F., Franzoni, M., Dongiovanni, D., Ballarini, M., Ferrari, S., D'Apolito, M., Di Cola, G., Capranico, G., Khobta, A., Campanini, R., Paolucci, P., Minucci, S. & Pession, A. 2006, 'G1 cell-cycle arrest and apoptosis by histone deacetylase inhibition in MLL-AF9 acute myeloid leukemia cells is p21 dependent and MLL-AF9 independent', *Leukemia*, vol. 20, no. 7, pp. 1307-10.

Trojanowska, M. 2009, 'Noncanonical transforming growth factor beta signaling in scleroderma fibrosis', *Curr Opin Rheumatol*, vol. 21, no. 6, pp. 623-9.

Tsujiuchi, T., Tsutsumi, M., Sasaki, Y., Takahama, M. & Konishi, Y. 1999, 'Hypomethylation of CpG sites and c-myc gene overexpression in hepatocellular carcinomas, but not hyperplastic nodules, induced by a choline-deficient L-amino acid-defined diet in rats', *Jpn J Cancer Res*, vol. 90, no. 9, pp. 909-13.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P. & Zhang, Y. 2006, 'Histone demethylation by a family of JmjC domain-containing proteins', *Nature*, vol. 439, no. 7078, pp. 811-6.

Uckun, F. M., Nachman, J. B., Sather, H. N., Sensel, M. G., Kraft, P., Steinherz, P. G., Lange, B., Hutchinson, R., Reaman, G. H., Gaynon, P. S. & Heerema, N. A. 1998a, 'Clinical significance of Philadelphia chromosome positive pediatric acute lymphoblastic leukemia in the context of contemporary intensive therapies: a report from the Children's Cancer Group', *Cancer*, vol. 83, no. 9, pp. 2030-9.

Uckun, F. M., Sensel, M. G., Sather, H. N., Gaynon, P. S., Arthur, D. C., Lange, B. J., Steinherz, P. G., Kraft, P., Hutchinson, R., Nachman, J. B., Reaman, G. H. & Heerema, N. A. 1998b, 'Clinical significance of translocation t(1;19) in childhood acute lymphoblastic leukemia in the context of contemporary therapies: a report from the Children's Cancer Group', *J Clin Oncol*, vol. 16, no. 2, pp. 527-35.

Ueki, T., Walter, K. M., Skinner, H., Jaffee, E., Hruban, R. H. & Goggins, M. 2002, 'Aberrant CpG island methylation in cancer cell lines arises in the primary cancers from which they were derived', *Oncogene*, vol. 21, no. 13, pp. 2114-7.

Van Beek, J. P., Kennedy, L., Rockel, J. S., Bernier, S. M. & Leask, A. 2006, 'The induction of CCN2 by TGFbeta1 involves Ets-1', *Arthritis Res Ther*, vol. 8, no. 2, p. R36.

van der Velden, V. H. & van Dongen, J. J. 2009, 'MRD detection in acute lymphoblastic leukemia patients using Ig/TCR gene rearrangements as targets for real-time quantitative PCR', *Methods Mol Biol*, vol. 538, pp. 115-50.

Van Dijk, M. A., Voorhoeve, P. M. & Murre, C. 1993, 'Pbx1 is converted into a transcriptional activator upon acquiring the N-terminal region of E2A in pre-B-cell acute lymphoblastoid leukemia', *Proc Natl Acad Sci U S A*, vol. 90, no. 13, pp. 6061-5.

Vaux, D. L., Cory, S. & Adams, J. M. 1988, 'Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells', *Nature*, vol. 335, no. 6189, pp. 440-2.

Volinia, S., Galasso, M., Costinean, S., Tagliavini, L., Gamberoni, G., Drusco, A., Marchesini, J., Mascellani, N., Sana, M. E., Abu Jarour, R., Desponts, C., Teitell, M., Baffa, R., Aqeilan, R., Iorio, M. V., Taccioli, C., Garzon, R., Di Leva, G., Fabbri, M., Catozzi, M., Previati, M., Ambs, S., Palumbo, T., Garofalo, M., Veronese, A., Bottoni, A., Gasparini, P., Harris, C. C., Visone, R., Pekarsky, Y., de la Chapelle, A., Bloomston, M., Dillhoff, M., Rassenti, L. Z., Kipps, T. J., Huebner, K., Pichiorri, F., Lenze, D., Cairo, S., Buendia, M. A., Pineau, P., Dejean, A., Zanesi, N., Rossi, S., Calin, G. A., Liu, C. G., Palatini, J., Negrini, M., Vecchione, A., Rosenberg, A. & Croce, C. M. 2010, 'Reprogramming of miRNA networks in cancer and leukemia', *Genome Res*, vol. 20, no. 5, pp. 589-99.

Vorwerk, P., Wex, H., Hohmann, B., Oh, Y., Rosenfeld, R. G. & Mittler, U. 2000, 'CTGF (IGFBP-rP2) is specifically expressed in malignant lymphoblasts of patients with acute lymphoblastic leukaemia (ALL)', *Br J Cancer*, vol. 83, no. 6, pp. 756-60.

Vucic, E. A., Brown, C. J. & Lam, W. L. 2008, 'Epigenetics of cancer progression', *Pharmacogenomics*, vol. 9, no. 2, pp. 215-34.

Wagner, W., Wein, F., Roderburg, C., Saffrich, R., Faber, A., Krause, U., Schubert, M., Benes, V., Eckstein, V., Maul, H. & Ho, A. D. 2007, 'Adhesion of hematopoietic progenitor cells to human mesenchymal stem cells as a model for cell-cell interaction', *Exp Hematol*, vol. 35, no. 2, pp. 314-25.

Wahab, N. A., Weston, B. S. & Mason, R. M. 2005, 'Connective tissue growth factor CCN2 interacts with and activates the tyrosine kinase receptor TrkA', *J Am Soc Nephrol*, vol. 16, no. 2, pp. 340-51.

Wahlfors, J., Hiltunen, H., Heinonen, K., Hamalainen, E., Alhonen, L. & Janne, J. 1992, 'Genomic hypomethylation in human chronic lymphocytic leukemia', *Blood*, vol. 80, no. 8, pp. 2074-80.

Wallis, J. P. & Reid, M. M. 1989, 'Bone marrow fibrosis in childhood acute lymphoblastic leukaemia', *J Clin Pathol*, vol. 42, no. 12, pp. 1253-4.

Wang, L. C., Swat, W., Fujiwara, Y., Davidson, L., Visvader, J., Kuo, F., Alt, F. W., Gilliland, D. G., Golub, T. R. & Orkin, S. H. 1998, 'The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow', *Genes Dev*, vol. 12, no. 15, pp. 2392-402.

Wang, Q., Usinger, W., Nichols, B., Gray, J., Xu, L., Seeley, T. W., Brenner, M., Guo, G., Zhang, W., Oliver, N., Lin, A. & Yeowell, D. 2011, 'Cooperative interaction of CTGF and TGF-beta in animal models of fibrotic disease', *Fibrogenesis Tissue Repair*, vol. 4, no. 1, p. 4.

Wang, Q., Williamson, M., Bott, S., Brookman-Amissah, N., Freeman, A., Nariculam, J., Hubank, M. J., Ahmed, A. & Masters, J. R. 2007, 'Hypomethylation of WNT5A, CRIP1 and S100P in prostate cancer', *Oncogene*, vol. 26, no. 45, pp. 6560-5.

Wang, Q., Zhang, A., Li, R., Liu, J., Xie, J., Deng, A., Feng, Y. & Zhu, Z. 2008, 'High glucose promotes the CTGF expression in human mesangial cells via serum and glucocorticoid-induced kinase 1 pathway', *J Huazhong Univ Sci Technolog Med Sci*, vol. 28, no. 5, pp. 508-12.

Watanabe, Y. & Maekawa, M. 2010, 'Methylation of DNA in cancer', *Adv Clin Chem*, vol. 52, pp. 145-67.

Weisberg, E., Wright, R. D., McMillin, D. W., Mitsiades, C., Ray, A., Barrett, R., Adamia, S., Stone, R., Galinsky, I., Kung, A. L. & Griffin, J. D. 2008, 'Stromal-mediated protection of tyrosine kinase inhibitor-treated BCR-ABL-expressing leukemia cells', *Mol Cancer Ther*, vol. 7, no. 5, pp. 1121-9.

Welner, R. S., Pelayo, R. & Kincaid, P. W. 2008, 'Evolving views on the genealogy of B cells', *Nat Rev Immunol*, vol. 8, no. 2, pp. 95-106.

Wenger, C., Ellenrieder, V., Alber, B., Lacher, U., Menke, A., Hameister, H., Wilda, M., Iwamura, T., Beger, H. G., Adler, G. & Gress, T. M. 1999, 'Expression and differential regulation of connective tissue growth factor in pancreatic cancer cells', *Oncogene*, vol. 18, no. 4, pp. 1073-80.

Whitlock, J. A., Sather, H. N., Gaynon, P., Robison, L. L., Wells, R. J., Trigg, M., Heerema, N. A. & Bhatia, S. 2005, 'Clinical characteristics and outcome of children with Down syndrome and acute lymphoblastic leukemia: a Children's Cancer Group study', *Blood*, vol. 106, no. 13, pp. 4043-9.

Whitman, S. P., Strout, M. P., Marcucci, G., Freud, A. G., Culley, L. L., Zeleznik-Le, N. J., Mrozek, K., Theil, K. S., Kees, U. R., Bloomfield, C. D. & Caligiuri, M. A. 2001, 'The partial nontandem duplication of the MLL (ALL1) gene is a novel rearrangement that generates three distinct fusion transcripts in B-cell acute lymphoblastic leukemia', *Cancer Res*, vol. 61, no. 1, pp. 59-63.

Wiemels, J., Kang, M. & Greaves, M. 2009, 'Backtracking of leukemic clones to birth', *Methods Mol Biol*, vol. 538, pp. 7-27.

Wilson, A., Laurenti, E. & Trumpp, A. 2009, 'Balancing dormant and self-renewing hematopoietic stem cells', *Curr Opin Genet Dev*, vol. 19, no. 5, pp. 461-8.

Wilson, A., Oser, G. M., Jaworski, M., Blanco-Bose, W. E., Laurenti, E., Adolphe, C., Essers, M. A., Macdonald, H. R. & Trumpp, A. 2007, 'Dormant and self-renewing hematopoietic stem cells and their niches', *Ann NY Acad Sci*, vol. 1106, pp. 64-75.

Wilson, A. & Trumpp, A. 2006, 'Bone-marrow haematopoietic-stem-cell niches', *Nat Rev Immunol*, vol. 6, no. 2, pp. 93-106.

Wood, M. A., McMahon, S. B. & Cole, M. D. 2000, 'An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc', *Mol Cell*, vol. 5, no. 2, pp. 321-30.

Xie, S., Sukkar, M. B., Issa, R., Oltmanns, U., Nicholson, A. G. & Chung, K. F. 2005, 'Regulation of TGF-beta 1-induced connective tissue growth factor expression in airway smooth muscle cells', *Am J Physiol Lung Cell Mol Physiol*, vol. 288, no. 1, pp. L68-76.

Xie, Y., Yin, T., Wiegand, W., He, X. C., Miller, D., Stark, D., Perko, K., Alexander, R., Schwartz, J., Grindley, J. C., Park, J., Haug, J. S., Wunderlich, J. P., Li, H., Zhang, S., Johnson, T., Feldman, R. A. & Li, L. 2009, 'Detection of functional haematopoietic stem cell niche using real-time imaging', *Nature*, vol. 457, no. 7225, pp. 97-101.

Xu, Y., Fang, F., Dhar, S. K., Bosch, A., St Clair, W. H., Kasarskis, E. J. & St Clair, D. K. 2008, 'Mutations in the SOD2 promoter reveal a molecular basis for an activating protein 2-dependent dysregulation of manganese superoxide dismutase expression in cancer cells', *Mol Cancer Res*, vol. 6, no. 12, pp. 1881-93.

Yagi, T., Hibi, S., Takanashi, M., Kano, G., Tabata, Y., Imamura, T., Inaba, T., Morimoto, A., Todo, S. & Imashuku, S. 2002, 'High frequency of Ikaros isoform 6 expression in acute myelomonocytic and monocytic leukemias: implications for up-regulation of the antiapoptotic protein Bcl-XL in leukemogenesis', *Blood*, vol. 99, no. 4, pp. 1350-5.

Yanagita, T., Kubota, S., Kawaki, H., Kawata, K., Kondo, S., Takano-Yamamoto, T., Tanaka, S. & Takigawa, M. 2007, 'Expression and physiological role of CCN4/Wnt-induced secreted protein 1 mRNA splicing variants in chondrocytes', *FEBS J* vol. 274, no. 7, pp. 1655-65.

Yang, D. H., Kim, H. S., Wilson, E. M., Rosenfeld, R. G. & Oh, Y. 1998, 'Identification of glycosylated 38-kDa connective tissue growth factor (IGFBP-related protein 2) and proteolytic fragments in human biological fluids, and up-regulation of IGFBP-rP2 expression by TGF-beta in Hs578T human breast cancer cells', *J Clin Endocrinol Metab*, vol. 83, no. 7, pp. 2593-6.

Yang, F., Tuxhorn, J. A., Ressler, S. J., McAlhany, S. J., Dang, T. D. & Rowley, D. R. 2005, 'Stromal expression of connective tissue growth factor promotes angiogenesis and prostate cancer tumorigenesis', *Cancer Res*, vol. 65, no. 19, pp. 8887-95.

Yang, M., Huang, H., Li, J., Li, D. & Wang, H. 2004, 'Tyrosine phosphorylation of the LDL receptor-related protein (LRP) and activation of the ERK pathway are required for connective tissue growth factor to potentiate myofibroblast differentiation', *FASEB J*, vol. 18, no. 15, pp. 1920-1.

Yi, F., Pereira, L. & Merrill, B. J. 2008, 'Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal', *Stem Cells*, vol. 26, no. 8, pp. 1951-60.

Yin, D., Chen, W., O'Kelly, J., Lu, D., Ham, M., Doan, N. B., Xie, D., Wang, C., Vadgama, J., Said, J. W., Black, K. L. & Koeffler, H. P. 2010, 'Connective tissue growth factor associated with oncogenic activities and drug resistance in glioblastoma multiforme', *Int J Cancer*, vol. 127, no. 10, pp. 2257-67.

Yoshida, N., Yagasaki, H., Xu, Y., Matsuda, K., Yoshimi, A., Takahashi, Y., Hama, A., Nishio, N., Muramatsu, H., Watanabe, N., Matsumoto, K., Kato, K., Ueyama, J., Inada, H., Goto, H., Yabe, M., Kudo, K., Mimaya, J., Kikuchi, A., Manabe, A., Koike, K. & Kojima, S. 2009, 'Correlation of clinical features with the mutational status of GM-CSF signaling pathway-related genes in juvenile myelomonocytic leukemia', *Pediatr Res*, vol. 65, no. 3, pp. 334-40.

Youlden, D., Baade, P., Ward, L., Valery, P., Hassell, T., Green, A. C. & Aitken, J. F. 2010, *Childhood cancer survival in Australia, 1995-2004*, Viertel Center for Research in Cancer Control, Brisbane, Queensland. Accessed: Jan 2010.

Yu, C. C., Hsu, M. J., Kuo, M. L., Chen, R. F., Chen, M. C., Bai, K. J., Yu, M. C., Chen, B. C. & Lin, C. H. 2009, 'Thrombin-induced connective tissue growth factor expression in human lung fibroblasts requires the ASK1/JNK/AP-1 pathway', *J Immunol*, vol. 182, no. 12, pp. 7916-27.

Zhang, Q., Ying, J., Zhang, K., Li, H., Ng, K. M., Zhao, Y., He, Q., Yang, X., Xin, D., Liao, S. K., Tao, Q. & Jin, J. 2007, 'Aberrant methylation of the 8p22 tumor suppressor gene DLC1 in renal cell carcinoma', *Cancer Lett*, vol. 249, no. 2, pp. 220-6.

Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., Yu, J., Lin, J. D., Wang, C. Y., Chinnaiyan, A. M., Lai, Z. C. & Guan, K. L. 2008, 'TEAD mediates YAP-dependent gene induction and growth control', *Genes Dev*, vol. 22, no. 14, pp. 1962-71.

Zivny, J., Klener, P., Jr., Pytlik, R. & Andera, L. 2010, 'The role of apoptosis in cancer development and treatment: focusing on the development and treatment of hematologic malignancies', *Curr Pharm Des*, vol. 16, no. 1, pp. 11-33.

Zumkeller, W. 2002, 'The insulin-like growth factor system in hematopoietic cells', *Leuk Lymphoma*, vol. 43, no. 3, pp. 487-91.