School of Biomedical Science

Insights into the alpaca skin transcriptome in relation to fibre colour

Rhys Cransberg

This thesis is presented for the Degree of

Doctor of Philosophy

Of

Curtin University

January 2017

Declaration of Authenticity

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

The research presented and reported in this thesis was conducted in compliance with the National Health and Medical Research Council Australian code for the care and use of animals for scientific purposes 8th edition (2013). The proposed research study received animal ethics approval from the Curtin University Animal Ethics Committee, Approval Number N43-09.

Signed 

Date 09/01/2017

Acknowledgements

Above all else I would like to thank my supervisor Dr. Kylie Munyard. Your unwavering belief and determination that I deserve to and should complete this thesis is the reason that it is here today. Thank you so much for your many years of guidance that have been so significant to my professional development.

I would also like to acknowledge the contributions of my co-supervisors Dr. Paula Moolhuijzen and Dr. Keith Gregg for the outstanding expertise they lent to this work. My extended thanks goes to the staff of the Schools of Biomedical Science and CHIRI many of whom have lent their knowledge over the duration of this project, in particular to Mrs Eleanor "Nell" Morgan whose Linux advice and assistance was vital.

My fellow research students and friends, of whom there are too many to list, thank you all for your friendship and support. In particular, I'd like to thank Ryan Harris for always being on the same page as me, Josh Ravensdale for always being agreeable, Kofi Stevens for his outstanding homebrew and Danielle Giustiniano for the excellent banter.

To my partner Emilie Loquier thank you for your patience, you have always been there for me when I've needed you.

I would also like to acknowledge the contribution of the Rural Industries Research and Development Corporation (RIRDC) who provided funding for the project.

Last but certainly not least I would like to thank my parents, who have shaped me into the person I am today. Thank you for your unconditional love and support, your belief and trust in me and also for the pride that you take with my achievements.

Abstract

Alpaca fibre products occupy a growing niche in the Australian textile industry. The two most economically important traits are the colour and the fineness of the fibre. Significant reductions in fibre diameter have been achieved, and genetic selection for fine fibre is well understood. However, the molecular basis of pigmentation in alpaca fibre is poorly understood. Breeders aiming to produce a specific fibre colour have little choice but to plan matings, or even purchase new breeding stock, with little to no knowledge of the underlying genetic basis of the colour. Subsequently, the desired colours are not always produced. Given the expense of high quality breeding stock, and a long gestation period (11-12 months), this unpredictability is a high cost for alpaca breeders, and the industry as a whole.

Colour in mammalian fibre is based on the combination and arrangement of two pigments, red to yellow pheomelanin, and black to brown eumelanin. The genes involved in the mammalian pigmentation pathway are highly conserved, which makes comparative genetics a useful tool in hypothesising the genes responsible for different colours in alpaca fibre. However, this comparison is complicated by the complexity of the pigmentation pathway – over 300 genes are involved in determining colour in mammals. Furthermore, all variants that have been shown to impact colour in alpaca fibre have been novel, with the exception of one recessive black agouti allele, which was first demonstrated in dogs. Understanding the types of pigments present in alpaca fibre through accurate, objective phenotyping is an essential component to finding the molecular causes of colour. Through a combination of spectrophotometric and HPLC methods I was able to demonstrate that brown coloured alpacas were predominantly pheomelanic, not eumelanic brown as is found, for example, in mice and chocolate Labradors. The distribution of pigment in brown alpacas is similar to bay coloration in horses, with eumelanic extremities and predominantly pheomelanic fibre on the body. Unexpectedly, the ratio of eumelanin to pheomelanin in the fibre of all bay (aka brown) and fawn phenotypes was similar, which indicated a role for colour dilution in alpaca phenotypes.

In order to gain a better understanding of the molecular basis of alpaca pigmentation, RNAseq was used to obtain a global view of all of the genes that were expressed in alpaca skin. In a proof-of-concept study to determine if RNAseq was sensitive enough to detect lowly expressed genes in a low abundance cell of the skin (i.e. melanocytes), polyA-RNA was extracted from small numbers of white, bay and black alpacas. The mRNA was pooled by colour, three Illumina TruSeq RNA sequencing libraries were prepared and barcoded, and then sequenced using an Illumina Genome Analyser II, which generated ~20 million paired end reads. These data showed that RNAseq was suitable for measuring the expression of colour determining genes in alpaca skin, despite the small percentage of melanocytes in the skin, and enabled the generation of the first annotation of the Vicugna_pacos_2.0.1 genome based on biological data in relation to skin.

Subsequently, the libraries prepared from the whole transcriptome of pooled RNA from 10 alpacas each for eight colours (white, light-skinned fawn, dark-skinned fawn, light bay, bay, roan, silvergrey and black) were sequenced using a SOLiD 5500xl sequencer, which generated ~1.25 billion paired end reads. All of the major colour determining genes expressed in the skin of other species were also expressed in alpaca skin. Furthermore, the expression of these genes followed a similar pattern to what would be expected in other mammals. For example, the expression of key genes in melanogenesis like *TYR*, *TYRP1*, *DCT*, *PMEL17* and *MART1* were all upregulated in more intensely pigmented animals such as bay and black, compared to the less intensely pigmented animals such as white and fawn. Similarly, the key antagonist to the production of eumelanin, *ASIP*, was found to be upregulated in lighter pigmented animals, especially white. The surprising discovery of a novel *ASIP* 5' non-coding exon, which mapped to the genome downstream and in the incorrect orientation, suggests that there could possibly be an inverted full or partial duplication of *ASIP* that is not yet captured in the reference genome. Similar duplications of *ASIP* occur in sheep and rabbits and are known to affect pigmentation. However confirmation of the presence, size and position of the *ASIP* duplication and how this might alter colour regulation was not able to be determined from the available data.

This is the first study to utilise RNAseq in alpacas. Most known mammalian pigmentation colour genes have never been analysed in alpacas from an expression perspective, and the global view of the expression profile of alpacas supports homology to other, more extensively studied mammals. As a result of this research alpaca breeders and researchers now have a clearer picture of the genes controlling colour, and the characterisation of the alpaca skin transcriptome has offered several new leads into the major genes affecting alpaca colour.

Table of Contents

List of Tables and Figures

List of Abbreviations

3-AHP	3-Amino 4-Hydroxyphenylalanine
4-AHP	4-Amino 3-Hydroxyphenylalanine
AAFL	Australian Alpaca Fleece Limited
α-MSH	Alpha Melanocyte Stimulating Hormone
ASIP	Agouti Signalling Protein
ATP	Adenosine Triphosphate
ATRN	Attractin
BAM	Binary Alignment Map
βDEF103	Beta-Defensin 103
BLAT	Blast Like Alignment Tool
bp, kb, Mb, Gb, Tb	Base Pairs, Kilobases, Megabases, Gigabases, Terabases
°C	Degrees Celsius
cAMP	Cyclic Adenosine Monophosphate
CCD	Charge Coupled Device
DAVID	Database for Annotation, Visualisation and Integrated Discovery
DCT	Dopachrome Tautomerase
DHI	Dihydroxyphenyaline
DHICA	Dihydroxyphenyaline Carboxilic Acid
DNA, cDNA	Deoxyribonucleic Acid, Complementary DNA
DOPA	Dihydroxyphenylalanine
EDN3	Endothelin-3
EDNRB	Endothelin Receptor Type B
FPKM	Fragments per Kilobase of Transcript per Million Reads Mapped
GFF	General Feature Format
GO	Gene Ontology
GWAS	Genome-Wide Association Study
HPLC	High Performance Liquid Chromatography
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KIT	Kit Oncogene
KITL	Kit Ligand
MART1	Melanoma Antigen Recognized by T Cells 1
MC1R	Melanocortin 1 Receptor
MGRN	Mahogutin
MITF	Microphthalmia-Associated Transcription Factor
MREG	Melanoregulin

MYO5A	Myosin-5a
NCOA6	Nuclear Receptor Coactivator 6
NGS	Next Generation Sequencing
OA1	Oculocutaneous Albinism 1
PAX3	Paired Box 3
PCR	Polymerase Chain Reaction
PDCA	Pyrole 2,3,5 Dicarboxylic Acid
PLDN	Palladin
PMEL17	Premelanosomal Protein
PTCA	Pyrole 2,3,5 Tricarboxylic Acid
RAB27A	Rab27a
RNA, mRNA, rRNA, miRNA ,snoRNA	Ribonucleic Acid, Messenger RNA, Ribosomal RNA, Micro RNA, Small Nucleolar RNA
RPKM	Reads Per Kilobase of Transcript Per Million Reads Mapped
s, mins	Seconds, Minutes
SABC	State Agricultural Biotechnology Centre
SAM	Sequence Alignment Map
SLC36A1	Solute Carrier Family 36 A1
SLC45A2	Membrane Associated Transport Protein
SNP	Single Nucleotide Polymorphism
SOX10	Sry-Box Containing Gene 10
TTCA	Thiazole 2,3,5-Tricarboxylic Acid
TYR	Tyrosinase
TYRP1	Tyrosinase Related Protein 1
VCF	Variant Call Format

1

Literature Review

1.1 Alpacas

1.1.1 History

Alpacas (originally *Lama pacos*, Linnaeus 1758; now *Vicugna pacos*, Wheeler 1995) played a central role in ancient Andean societies where they were traditionally bred for both meat and fibre (Mengoni Gonalons 2008). They are now prevalent as domesticated animals in parts of North and South America, Europe and Australia (Frank *et al.* 2006). They are members of the family Camelidae, which also includes camels (*Camelus*, Linneaus, 1758), llamas (*Lama glama*, Linneaus, 1758), vicunas (*Vicugna vicugna*, Molina 1782) and guanacos (*Lama guanicoe*, Müller 1776). Until recently alpacas were classified under the genus *Lama* (Cecchi *et al.* 2004), but evidence presented by Kadwell et al. (2001) suggested that this classification was incorrect and that alpacas were actually descendants of the vicuna, and hence they were placed in the genus *Vicugna* in 2001.

Alpacas held a place in religious beliefs for the Incan civilisation (Wheeler 1995; Mengoni Gonalons 2008). Uniformly-coloured alpacas of various colours were used in ritual sacrifice to specific deities, and herds were bred to meet this demand (Wheeler 1995). Some of the alpacas from around this time showed extraordinary fineness in their fibre; a characteristic which is also much sought after in modern times (Wheeler 1995). The Spanish conquest in the 16th century caused a major disruption to the alpacas of South America (Mengoni Gonalons 2008). During the first century after the Spanish conquest, around 90% of the South American camelids died through slaughter, disease, and competition with introduced species (Wheeler 1995). This created a population bottleneck, while changes in animal management led to extensive hybridisation with llamas (Wheeler 1995; Wheeler *et al.* 1995; Kadwell *et al.* 2001). This hybridisation and reduction of genetic diversity resulted in the diminution of the extra-fine fibre trait, increased hairiness and both increased coarseness and variation in fibre diameter observed in modern huacaya and suri alpacas. These desirable traits are currently being selected for, and managed in better-quality breeding programs (Wheeler *et al.* 1995).

1.1.2 The Alpaca Industry

In 2006 there were approximately 3.5 million alpacas worldwide; the vast majority - over 90%, in South America (Lupton *et al.* 2006). Alpaca fibre is highly valued because it is strong, soft and lustrous, and suitable for use in a wide range of fabrics from high quality clothing to carpets (Lupton *et al.* 2006; McGregor 2006; Davison 2008). Even relatively course alpaca fibre adds to the value of carpets due to its exceptional softness. Further, alpaca fibre offers a wide selection of natural colours which are unavailable among more conventional natural fibres such as sheep wool and cotton (Lupton *et al.* 2006).

Peru is the largest producer of alpaca fibre, though there have been significant increases in production in the Australian alpaca industry (McGregor 2006). In 2012 the registered population of alpacas in Australia was over 150,000 animals (Australian Alpaca Fleece Limited, 2012). This is an increase from the estimated 100,000 animals registered in 2008 (Davison 2008) highlighting the growth of the Australian alpaca industry. In 2007, Australian alpacas produced 70 tonnes of fibre, making Australia the largest producer of fibre outside of South America (Davison 2008). Other countries with developing alpaca industries include the United States (170,000 alpacas registered in 2012), United Kingdom (approximately 20,000), Canada (12,000 registered in 2012), New Zealand (10,000), Germany (7,000; Davison, 2008) and South Africa (1,200).

1.1.3 Current Limitations for the Australian Alpaca Industry

A range of factors, including fibre type, length, uniformity of fibre diameter and colour, all influence the price of alpaca fibre (Frank *et al.* 2006). White fibre is considered the most valuable because it is easy to dye, and in Australia coloured fibre is less valuable than white fibre (Table 1.1).

However, there are not enough white alpacas in Australia to meet market demand, with only approximately a third of Australian alpacas being light enough to be classed as white (Table 1.2). This is in contrast with established alpaca industries such as that in Peru, where 50 – 56% of alpacas are white (Frank *et al.* 2006). There

Table 1.1. Prices are A$/kg of whole shorn fibre including neck and pieces, GST excluded (Australian Alpaca Fleece Limited 2012).

	X-Fine	Fine	Medium	Adult
White	3.57	2.96	1.51	0.58
Colour	2.55	1.93	0.26	0.10

is a recent trend for more naturally coloured alpaca fibre to be used by commercial alpaca breeders and organisations, taking advantage of the natural colours which are not available in sheep wool.

Table 1.2. The number and proportions of major alpaca colour types in Australia, (Australian Alpaca Fleece Limited 2012).

Colour	Number of Alpacas	% Australian Population
White	49776	33.2%
Light Fawn	19383	12.9%
Black	19121	12.7%
Mid Fawn	12903	8.5%
Dark Fawn	9921	6.6%
Mid Brown	9436	6.3%
Two or More Colours	5199	3.4%
No Colour Listed	405	0.3%

To enable breeders to produce enough fibre of a specific colour to meet market demand, it is essential to understand the underlying colour genetics of the species, i.e. how the different colours are inherited. The molecular mechanisms for determining fibre colour in alpacas are beginning to be elucidated (Powell *et al.*

2008; Feeley & Munyard 2009; Feeley *et al.* 2011; Chandramohan *et al.* 2013; Feeley *et al.* 2016). Information on mutations that influence fibre colours such as those already identified in the *melanocortin-1 receptor* (Feeley & Munyard 2009) and *agouti signalling protein* (Feeley *et al.* 2011) enable breeders to plan mating to more effectively achieve the desired fibre colours, and to avoid unwanted colours.

1.1.4 Types of Alpaca Fibre Colour

The nomenclature used to describe colour in alpaca is inconsistent, and highly subjective. In order to standardise these phenotype descriptions, guidelines were formed in our research group based on assignment criteria which focus on the predominant melanin type observed in each colour (Cransberg *et al.* 2013; Table 1.3). The most notable change is that alpacas that would have been presented and described by alpaca breeders and organisations as brown are classed in this document as bay, as this phenotype much more closely resembles the bay colour in horses with a red coat and dark points as opposed to a "true" brown colour, such as seen in brown mice and chocolate Labradors.

1.1.5 Inheritance of Colour in Alpacas

Until 2008 there had been no publications on the molecular genetic basis of colour in alpacas. Alpaca fibre research had been directed toward fibre quality due to the financial significance of that characteristic (Frank *et al.* 2006; McGregor 2006). The challenge of understanding alpaca colour genetics is exemplified by the plethora of shades of alpaca fibre colour, making it highly subjective to know which colour phenotype to assign to any given individual (Paul 2006). It is thought that alpaca fibre colour is caused by complex interactions between genes at different loci, similar to that seen in many other mammalian species. Pigmentation in mammals involves many genes which have been demonstrated to act at several stages – from the development and migration of melanocytes, to the types and amount of pigments that are produced, to how the mature melanocytes are located; all of which can result in a difference in pigmentation.

Table 1.3. Guidelines used to assign animal samples to a colour group for the colours analysed. Adapted from Cransberg *et al.* (2013).

Phenotype description	**Colour Phenotype**
Very weakly or un-pigmented fibre with pink skin	White
Pale to moderate intensity yellow to red coloured fibre, with pink skin	Pink-skinned Fawn
Pale to moderate yellow to red coloured fibre, with dark skin	Dark-skinned Fawn
Black extremities, pale yellow to pale red coloured fibre and dark skin	Light Bay
Black extremities, yellow to dark red coloured fibre and dark skin	Bay
eumelanic fibre, diluted over whole body, white fibre on face, with or without non-diluted patches with dark skin	Silver-grey
Black fibre mixed with white fibre and dark skin	Roan
Dark skin and black fibre	Black

Several studies into the molecular basis of alpaca fibre colour have focused on the *MC1R* gene. Feeley and Munyard (2009) demonstrated two mutations, which could potentially completely eliminate the function of the MC1R protein, which in turn will lead to a phenotype where the animal cannot produce eumelanin. This phenotype results in light or pink skinned individuals, with no eumelanin in the fibre. Powell *et al.* (2008) described the same mutations, but concluded that they did not cause a change in phenotype because they did not correlate to a single colour.

A report by Munyard (2011) suggested Mendelian patterns of inheritance for some alpaca fibre colours, based on breeding records, comparative genetics and research conducted on candidate alpaca genes (Powell *et al.* 2008; Feeley & Munyard 2009; Cransberg & Munyard 2011; Feeley *et al.* 2011). In this report it is proposed that the *MC1R* and *ASIP* genes play key roles in determining the base fibre colour in alpacas.

Further studies have begun to focus on changes in candidate gene expression levels (Chandramohan *et al.* 2013; Tian *et al.* 2015) and miRNAs (Yang *et al.* 2015).

1.2 Pigmentation

1.2.1 A Brief History On the Study of Mammalian Pigmentation

Mammalian pigmentation has interested people for many centuries. One of the earliest written references to pigmentation was thought to have been in the Aushooran period, around 2200BC, and was thought to describe vitiligo (Najamabadi 1934).

The Roman scholar Pliny (23-79 AD) once remarked on the oddities of human pigmentation:

"One certain example is that of the renowned boxer Nicaeus, born at Byzantium, whose mother was the daughter of adultery with a Negro. Her complexion was no different from that of the others [other white women], but her son Nicaeus appeared like his Negro grandfather." –Naturalis Historia VII.12.51

This demonstrates one of the early-recorded examples of the curiosity that has inspired research into the causes of varying levels of pigmentation in humans. Throughout the course of scientific advancement, the differences in dark to light human skin has been investigated. In 1716 Albinus published a paper with an illustration of dissected black skin (Westerhof 2006). Some of the earliest microscopic and chemical analysis of melanins - known only as "pigment" at the time - were conducted to compare dark and light skin (Abel & Davis 1896). These authors reported that it was likely that the type of pigment was similar between light and dark skin, which later proved to be accurate. They also described the chemical properties of "pigments", many of which, such as the solubility and insolubility of pigment molecules under various acid and alkaline conditions are the basis of chemical melanin characterisation today (Abel & Davis 1896).

The discovery of pigment producing cells – later known as melanocytes - is accredited to two scientists: Henle (1837) and Simon (1841), who independently reported the discovery of these cells within four years of each other. It was noted that these cells were responsible for the creation of the "pigment" which was present in human skin. Only later were the names eumelanin and pheomelanin coined and defined (Görnitz 1923).

Pigmentation is a topic that is relatively straightforward to study. Changes in genotype will normally manifest in a variation of the colour in question; depending on the subtlety of the phenotype change, and can often be confirmed visually. It was no surprise that the first recognition of the occurrence of Mendelian traits in mammals were found in studies of pigmentation by Cuénot (1902) who demonstrated Mendelian ratios of inheritance within different coloured mice. Soon after, similar patterns of Mendelian inheritance were discovered in rabbits (Hurst 1905) and horses (Hurst 1906). The pattern of colour inheritance in domesticated animals was only important at this stage as a means of determining breed purity, and as a "source of scientific entertainment" (Simpson 1914).

In 1989 one of the first known colour-determining mutations was characterised (Tomita *et al.* 1989). This was a mutation in the human *TYR* gene, which is the cause of type 1 oculocutaneous albinism (OCA1). Over 303 different mutations which result in albinism have now been found in the *TYR* gene (Mártinez-García & Montoliu 2013). With the advancement of molecular genetic technologies, polymorphisms affecting colour were discovered at over 50 loci in mice by 1979. By the early 2000s this number had increased to over 100 different loci (Bennett & Lamoreux 2003) and it continues to increase, until by 2016, 171 different genes have been demonstrated, via cloning into mice, to affect hair colour (http://www.espcr.org/micemut/).

1.2.2 Pigmentation

1.2.2.1 Melanocytes and melanosomes

Melanocytes are pigment producing cells that exist in the skin, eyes, inner-ear, mucus membranes, meninges and central nervous system of mammals (Sulaimon &

Kitchell 2003; Tolleson 2005). Types of melanocytes which exist in mammals include cutaneous, ocular, otic and cephalic melanocytes (Tolleson 2005). The study of pigmentation is primarily concerned with cutaneous melanocytes. Melanocytes in different parts of the body produce melanin with different properties and functions. For example, melanin protects the eyes and brain by trapping metals and free radicals (Korytowski *et al.* 1995; Zaręba *et al.* 1995; Nicolaus 2005).

1.2.2.2 Melanin

Melanin is a mixture of complex biopolymers, the concentrations and arrangement of which determine the colour of skin and hair in mammals (Ito & Wakamatsu 2003; Hoekstra 2006). Two types of melanin can be produced via the melanin synthesis pathway in mammals: eumelanin; a black to brown pigment with ellipsoidal granules; and pheomelanin; a yellow to red spherical pigment granule (Ito & Wakamatsu 2003; Hoekstra 2006). It is the amount and pattern of these molecules which result in the spectrum of colours observed in mammalian pigmentation (Barsh 2006). Eumelanin and pheomelanin share a similar arrangement of molecules linked by carbon-carbon bonds and are both derived from a common precursor, dopaquinone (Slominski *et al.* 2004; Ito & Wakamatsu 2008). However, they differ in their chemical composition and their structural and physical properties (Slominski *et al.* 2004; Tobin 2011). Two copolymers, dihydroxyindole (DHI) and dihydroxyindole carboxylic acid (DHICA) make up eumelanin molecules (Ozeki *et al.* 1997b). Pheomelanin is comprised of benzothiazine units, which are polymerised in the final stages of pheomelanogenesis.

The melanin synthesis pathway has been demonstrated in a number of studies to be highly conserved among mammals (Martinez-Arias *et al.* 2000; Camacho-Hübner *et al.* 2002; Wehrle-Haller 2003; Guibert *et al.* 2004; Schmutz *et al.* 2004; Wang & Hebert 2006; Deng *et al.* 2007). In fact, all three of the tyrosinase gene family proteins, the key enzymes of melanin synthesis in mammals, have homologous counterparts in amphibians and fish, suggesting conservation of the melanin pathway throughout the evolution of vertebrates (Camacho-Hübner *et al.* 2002; Kumasaka *et al.* 2003). Thus, we can compare the pigmentation pathways of other mammals to those of alpacas using comparative genetics.

1.2.3 Major Pigmentation Genes

1.2.3.1 Major Pigmentation Genes

The many genes shown to impact on colour can be loosely grouped into those effecting melanocyte migration and development; those effecting melanocyte formation and maturation, those involved with melanin synthesis, and those which effect the export of melanosomes from the melanocyte (Slominski *et al.* 2004; Yamaguchi & Hearing 2009; Figure 1.1; Table 1.4).

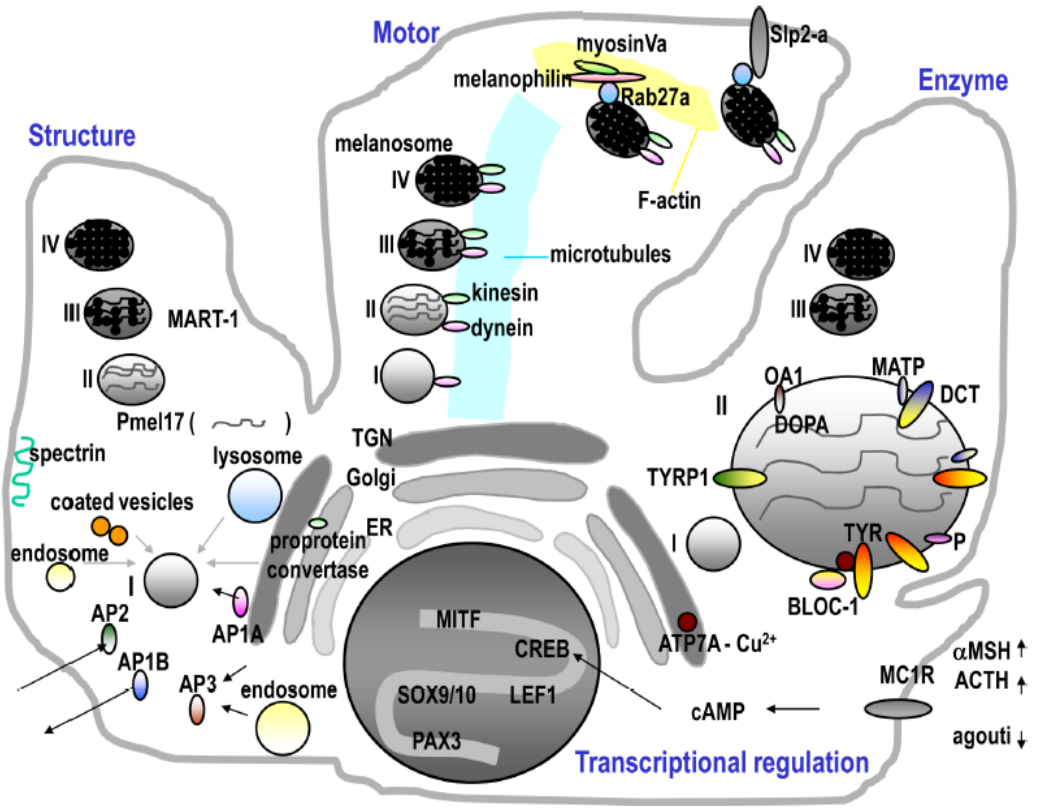

Figure 1.1. The major factors which influence pigmentation illustrated within a melanocyte. Adapted from Yamaguchi and Hearing (2009).

From all of the genes known to affect pigmentation in mammals a set were selected as the most likely to have an effect on pigmentation in alpacas, based on their role in pigmentation in other animals (Table 1.4).

Table 1.4. Pigmentation genes discussed and the abbreviations used.

Gene	Acronym	Process
Endothelin 3	*EDN-3*	Melanocyte Migration and Development
Endothelin Receptor Type B	*EDNRB*	
Kit Oncogene	*KIT*	
Kit Ligand	*KITL*	
Microphthalmia-Associated Transcription Factor	*MITF*	
Paired Box 3	*PAX-3*	
Sry-Box Containing Gene 10	*SOX-10*	
Agouti Signalling Protein	*ASIP*	Melanocyte Formation and Maturation
Attractin	*ATRN*	
B-Defensin 103	*B-DEF 103*	
Melanoma Antigen Recognized by T Cells 1	*MART1*	
Melanocortin 1 Receptor	*MC1R*	
Mahogutin	*MGRN*	
Palladin	*PLDN*	
Premelanosomal Protein	*PMEL17*	
Dopachrome Tautomerase	*DCT*	Melanin Synthesis
Membrane Associated Transport Protein	*SLC45A2*	
Solute Carrier Family 36 A1	*SLC36A1*	
Tyrosinase	*TYR*	
Tyrosinase Related Protein 1	*TYRP1*	
Melanoregulin	*MREG*	Export of Melanosomes
Myosin-5a	*MYO5A*	
Rab27a	*RAB27A*	

1.2.3.2 Migration and Melanocyte Development

In mammals and birds the precursors of melanocytes, neural crest derived melanoblasts, migrate laterally to the basal layer of the epidermis where they develop into mature melanocytes (Rawles 1947; Mayer 1973; Hoekstra 2006). A number of genes are involved in the survival and migration of melanoblasts, including the *kit*

oncogene (*KIT*) and *kit ligand* (*KITL*), *paired box 3* (*PAX-3*), *SRY-box containing gene 10* (*SOX-10*), *microphthalmia-associated transcription factor* (*MITF*), *endothelin receptor type B* (*EDNRB*) and *endothelin 3* (*EDN-3).*

KIT is a receptor tyrosine kinase, which is activated by its ligand: KITL. KIT activation results in the formation of KIT dimers, and the subsequent internalisation of the KIT-KITL complex leads to induction of multiple signalling pathways (Slominski *et al.* 2004). These combined signals drive the migration of the melanoblasts from the neural crest, and assists their survival; while melanoblasts that fail to migrate are destroyed (Wehrle-Haller & Weston 1999). Microphthalmia-associated transcription factor (MITF) activates *KIT* expression, which results in a multiplier effect (Hou *et al.* 2000; Cooper & Raible 2009). Increased levels of activated KIT stimulate the rate of MITF phosphorylation, which in turn leads to increased transcription of *MITF*, creating a cyclic process (Hou *et al.* 2000).

KITL and EDN-3 have been demonstrated to work cooperatively and indispensably in the melanocyte differentiation pathway (Aoki *et al.* 2005). The binding of EDN-3 to EDNRB activates signalling, resulting in melanocyte differentiation. In mice without activated EDNRB, no mature melanocytes are observed (Hou *et al.* 2000; Aoki *et al.* 2005). MITF has been shown to regulate the expression of *EDNRB*; and EDN-3, along with Endothelin-1, triggers the phosphorylation of *MITF*, enhancing its expression (Sato-Jin *et al.* 2008). This complex, positive-feedback interaction results in the upregulation of expression of these key genes, which in turn drives melanocyte differentiation. Furthermore, it has been suggested that *EDNRB* signalling is required for *KITL* expression, because addition of an external supply of KITL to EDNRB deficient cells allows melanoblasts to form into mature melanocytes (Hou *et al.* 2000).

PAX-3 and SOX-10 are transcription factors that bind to the *MITF-M* promoter (Saito *et al.* 2003a; Yang *et al.* 2008). Expression of either *PAX-3* or *SOX-10* is enough to switch on the expression of *MITF* (Potterf *et al.* 2000). However, when PAX-3 and SOX-10 act together synergism is observed, resulting in a 100-fold increase in *MITF* promoter activity (Sato *et al.* 1999; Potterf *et al.* 2000; Widlund & Fisher 2003). SOX-10 has also been demonstrated to directly upregulate expression

of one of the key enzymes of melanogenesis, dopachrome tautomerase (DCT) in a human promoter-reporter construct (Potterf *et al.* 2001). And it has been shown that SOX-10 can also work synergistically with MITF to activate the *DCT* gene (Ludwig *et al.* 2004).

MITF acts as a transcription factor for a number of processes in pigment production, from the development of pigment cells (Hou *et al.* 2000; Saito *et al.* 2003; Widlund & Fisher 2003) to regulating genes involved in the production of pigment (Aksan & Goding 1998). It belongs to a family of transcription factors that possess a basic helix-loop-helix and leucine zipper structure (Lister *et al.* 1999; Saito *et al.* 2003). There are at least 6 different reported isoforms of *MITF*, all of which have unique 5' ends, caused by multiple alternative promoters (Widlund & Fisher 2003). These isoforms all contain a similar carboxy terminus, containing the basic histidine-loop-histidine leucine zipper structure, as well as a serine rich region and a transcriptional activation domain (Slominski *et al.* 2004). *MITF* is regulated by environmental factors, such as UV light (Miyamura *et al.* 2007), which increases *MITF* expression, as well as factors that are secreted from keratinocytes, fibroblasts and other cells that include but are not limited to, the aforementioned PAX-3 and SOX-10 (Kondo & Hearing 2011). The *MITF-M* isoform is melanocyte specific, due to a melanocyte-restricted promoter enhancer (Yajima *et al.* 1999). The insertion of a LINE retrotransposable element into intron 3 of the *MITF-M* gene is the cause of the black eyed white phenotype in mice. The phenotype results from the loss of the melanocyte population from the epidermis but not the eyes (Yajima *et al.* 1999).

1.2.3.3 Melanosome Formation and Maturation

Once melanoblasts have matured into melanocytes, they produce melanosomes, vesicle like structures in which melanin is synthesised (Marks & Seabra 2001). The maturation of melanosomes can be summarised into four distinct stages (I-IV), based on the development of filaments and the amount of melanisation present (Sitaram & Marks 2012). Two types of melanosomes are produced: those producing eumelanin termed eumelanosomes, and those which produce pheomelanin - pheomelanosomes (Sulaimon & Kitchell 2003). The maturation stages are similar between eumelanosomes and pheomelanosomes, with the exception that pheomelanosomes remain circular through maturation whereas eumelanosomes become ellipsoidal

(Costin & Hearing 2007). The change from circular vesicle to ellipsoidal eumelanosomes is triggered by cleavage of melanogenesis related proteins, and this cleavage also causes the formation of an intramelanosomal fibrillar network, which allows for the highly structured, organised packaging of eumelanin (Slominski *et al.* 2004). Early stage melanosomes (stages I and II) do not produce significant amounts of melanin, but intraluminal fibril formation is more developed in stage II melanocytes than in stage I melanocytes (Yamaguchi & Hearing 2009). Late stage melanosomes (III and IV) are generally thought to share the same type of fibrillar structure – with the major difference being the intensity of melanisation. Stage IV melanosomes become so highly pigmented that internal structures are obscured by the high levels of melanin (Hoashi *et al.* 2005). Pheomelanosomes do not share the same intraluminal fibril formation, instead, pheomelanin is deposited in the pheomelanosome in a vesiculoglobular matrix (Jimbow *et al.* 2000). The genes, which drive the maturation of pheomelanosomes, are not well described when compared to those of eumelanosomes.

Premelanosomal protein 17 (*PMEL17*) and *Melanoma Antigen Recognized by T cells 1* (*MART-1*) are two genes known to be critical to the maturation of eumelanosomes (Hoashi *et al.* 2005). PMEL17 is crucial in initiating the synthesis of intraluminal fibrils in eumelanosomes (Berson *et al.* 2001). Specifically, PMEL17 cleavage and processing accompanies the restructuring leading to maturation of stage I to stage II melanosomes (Kushimoto *et al.* 2001). Mutations in the *PMEL17* gene are responsible for the silver coat colour in mice (Kwon *et al.* 1995) and in ponies (Reissmann *et al.* 2007). PMEL17 and MART-1 have been demonstrated to form a complex in an early subcellular compartment, either in the golgi and/or the endoplasmic reticulum (Hoashi *et al.* 2005). Furthermore Hoashi *et al.* (2005) demonstrated that MART-1 regulated the expression, stability, trafficking, and processing of PMEL17, suggesting that it is essential for PMEL17 function, and thus for melanosomal maturation. Though there are no known *MART-1* mutations which cause a colour change, knockout mouse models have demonstrated that its removal results in the dilution of eumelanin in mice (Aydin *et al.* 2012). Stage III and IV melanosomes in the *MART-1* knockout mice have gaps between the pigmented compartment and the limiting membrane, while early stage melanosomes are unaffected, suggesting a role for MART1 in the production of late stage

melanosomes (Aydin *et al.* 2012). Both *PMEL17* and *MART-1* are transcriptionally regulated by MITF (Du *et al.* 2003).

The oculocutaneous albinism type 1 gene product (OA1) has more recently been identified as a key element in melanogenesis (Cortese *et al.* 2005). Unlike other forms of albinism, OA1 caused albinism results from a reduced number of melanosomes (Cortese *et al.* 2005), or through the formation of "giant" melanosomes termed macro-melanosomes (Incerti *et al.* 2000). It has been shown that in the absence of functional MART-1 the expression of *OA1* is severely depleted (Giordano *et al.* 2009), though this reaction is independent of the MART-1/PMEL17 interaction (Hoashi *et al.* 2005; Giordano *et al.* 2009).

The interactions between the melanocortin-1 receptor (MC1R) and agouti signaling protein (ASIP) determine whether eumelanosomes or pheomelanosomes will be produced by the melanocyte (Oyehaug *et al.* 2002). There are three possible types of alleles of the *MC1R* gene. Dominant alleles are generally associated with exclusively eumelanin production as seen, for example, in European black pigs (Kijas *et al.* 1998). Recessive alleles are often associated with pheomelanin production as seen in chestnut horses (Marklund *et al.* 1996), and the wild-type allele is neutral so expression from the *ASIP* locus will determine phenotype, and both pheomelanin and eumelanin can be produced (Furumura *et al.* 1996).

Traditionally, gain of function *ASIP* mutations have been associated with increased pheomelanin levels. More recently however, it has been demonstrated that increases in agouti levels correlate with decreases in levels of both eumelanin and pheomelanin in human melanocyte cultures (Le Pape *et al.* 2008). This occurs through the downregulation of many of the key genes involved in pigmentation (Le Pape *et al.* 2009). In contrast to MC1R, loss of function mutations in ASIP are associated with increased eumelanin production, demonstrated in agouti black rabbits, dogs, horses and alpacas (Rieder *et al.* 2001; Schmutz & Berryere 2007; Fontanesi *et al.* 2010; Feeley *et al.* 2011). Changes in the structure of the *ASIP* gene have been linked with switches between eumelanin and pheomelanin in alpacas, however it has also been shown that its expression can play a role in dilution of colour in other mammals. White coat colour in sheep is caused by a gene duplication of *ASIP* that brings it

under the control of a constitutively expressing promoter (Norris & Whan 2008). This leads to increased expression of the *ASIP* gene, which results in a phenotype with fewer mature melanocytes in the skin than in animals with lower expression of *ASIP* (Norris & Whan 2008). An inverted duplication in *ASIP* in mice is thought to be responsible for the *light-bellied agouti* (A^w) allele in mice (Chen *et al.* 1996). Either one of the two isoforms resulting from the duplication are able to be expressed, which can lead to the lighter ventral pigmentation observed (Hustad *et al.* 1995; Chen *et al.* 1996). A study on mainland and beach mice, which share a similar lighter ventral pattern phenotype to A^w mice, demonstrated that decreased expression of *ASIP* on the ventral side during embryogenesis resulted in fewer mature melanocytes when compared to the dorsal side, which had decreased *ASIP* expression (Manceau *et al.* 2011). *ASIP* has been shown to have multiple promoters and alternate 5'UTR sequences; with four alternate sequences identified in rabbits (Fontanesi *et al.* 2010), three in cows (Girardot *et al.* 2005), four in mice (Vrieling *et al.* 1994) and five in pigs (Drögemüller *et al.* 2006) with many of these transcripts containing multiple regulatory exons. In rabbits, mice and pigs alternate 5'UTR transcripts were differentially expressed between the dorsal and ventral regions (Vrieling *et al.* 1994; Drögemüller *et al.* 2006; Fontanesi *et al.* 2010).

Alpha-melanocyte stimulating hormone (α-MSH) is the antagonist to ASIP activity in relation to MC1R, and when it binds to MC1R eumelanin production is favoured (Busca & Ballotti 2000; Oyehaug *et al.* 2002). The binding of α-MSH to MC1R causes increased cAMP levels, which leads to increased eumelanin production (Widlund & Fisher 2003). This occurs because increased cAMP levels upregulate downstream signalling, most notably the *MITF* gene, which results in increased expression of several genes necessary for eumelanin production, including *TYR*, *TYRP1* and *DCT* (Busca & Ballotti 2000). Unlike most colour genes which are produced in melanocytes, α-MSH is produced in the anterior pituitary gland (Slominski *et al.* 2004).

Mutations in *mahogutin* (*MGRN*) and *attractin* (*ATRN*) will cause almost identical phenotypic effects, resulting in darker colouration of the fibre due to increased levels of pheomelanin (He *et al.* 2003). These genes are both thought to be accessory proteins to ASIP. Therefore mutations which decrease the activity in either

mahogutin or attractin will cause loss of ASIP function and consequently a eumelanic phenotype (He *et al.* 2003; Bagher *et al.* 2006).

The MC1R/ASIP interaction determines the major type of melanin produced, but there are many other genes that influence melanin production (Figure 1.2).

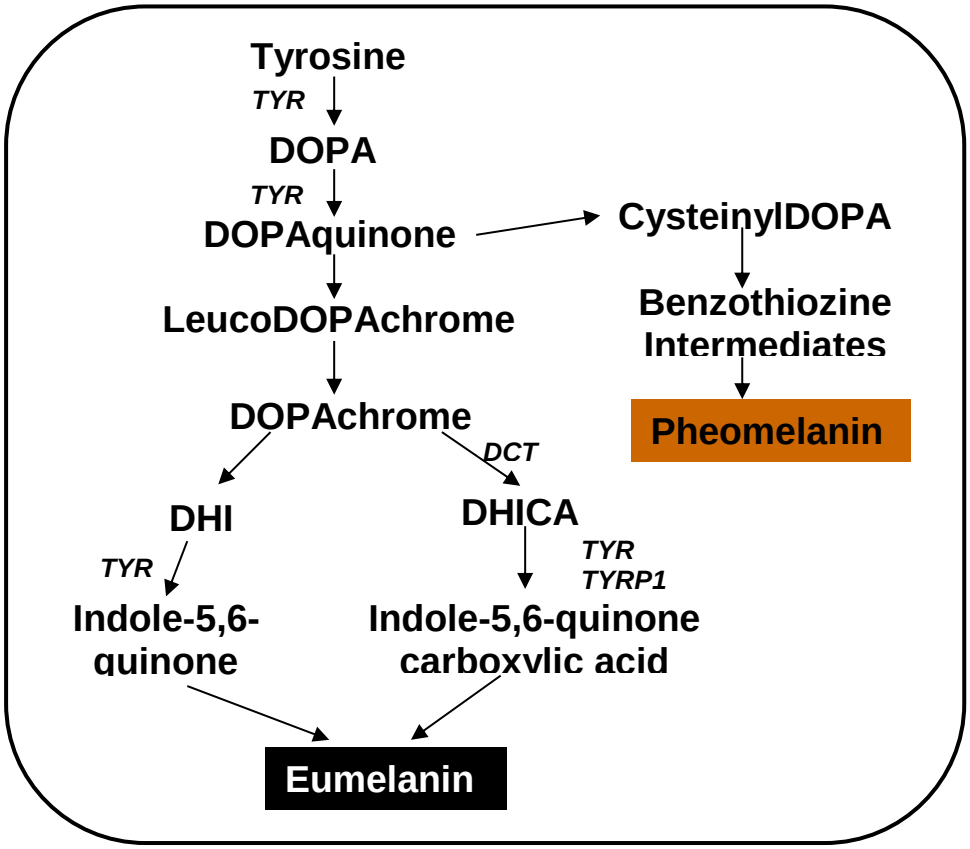

Figure 1.2 The melanin synthesis pathway: adapted from Munyard (2008) Key: TYR (tyrosinase); DCT (dopachrome tautomerase or tyrosinase related protein 2); TYRP1 (tyrosinase related protein 1).

1.2.3.4 Melanin Synthesis

The tyrosinase family of proteins are key to the conversion of L-tyrosine to either pheomelanin or eumelanin (Ito 2003). Tyrosinase (TYR) is involved in the initial reactions of the melanin synthesis pathway, and mutation to its structure is associated with albinism and colour dilution (Sulaimon & Kitchell 2003; Hirobe *et al.* 2006; Deng *et al.* 2007). TYR is a copper dependent metalloenzyme that is crucial to melanogenesis (Sulaimon & Kitchell 2003). TYR catalyses three key reactions in the melanin synthesis pathway, firstly the oxidation of tyrosine in the production of

dihydroxyphenylaline (DOPA) and the subsequent oxidation of DOPA to dopaquinone (Mason *et al.* 1955; Korner & Pawelek 1982; Figure 1.2). Dopaquinone is a common intermediate in eumelanin and pheomelanin production, without it neither of these two pigmentation pathways can function properly (Jimbow *et al.* 1991). The third step is through the oxidation of DHI in the production of DHI-eumelanin and/or the oxidation of DHICA in DHICA-eumelanin production (Tsukamoto *et al.* 1992).

The second tyrosinase family enzyme is tyrosinase related protein 1 (TYRP1). The major function of TYRP1 in mice is the oxidation of DHICA (Kobayashi *et al.* 1994b). However, it has been demonstrated that this function does not occur in humans (Boissy *et al.* 1998). Additionally TYRP1 is also known to stabilise TYR in mice, as TYR is more quickly degraded in melanocytes without fully functional TYRP1 (Kobayashi *et al.* 1998). In mice it was demonstrated that in the absence of functional TYRP1, PMEL17 might also promote DHICA oxidation, to allow efficient eumelanin production (Chakraborty *et al.* 1996). The absence of a functional TYRP1 protein can result in a brown phenotype (Bennett 1991), as seen in *TYRP1* knockout mice and chocolate Labradors (Zdarsky *et al.* 1990; Schmutz *et al.* 2002). Given that TYRP1 has a role in the oxidation of DHICA-eumelanin and no role in the formation of DHI-eumelanin, the absence of a functional TYRP1 protein might be expected to create an effect on the ratio between DHICA-eumelanin and DHI-eumelanin that could result in brown eumelanin rather than black eumelanin (Ozeki *et al.* 1995; Wilczek *et al.* 1996). However this is not the case. The difference between black and brown eumelanosomes is in the smaller molecular size of eumelanin molecules in brown skin/hair, which are present in round and relatively disorganized brown eumelanosomes as opposed to larger eumelanin molecules in ellipsoidal black eumelanosomes (Ozeki *et al.* 1997b; Kobayashi *et al.* 1998).

The third enzyme in the TYR family, (DCT) otherwise known as tyrosinase related protein 2, is known to catalyse the tautomerisation of dopachrome in the production of DHICA (Sulaimon & Kitchell 2003). A mutation causing a change from an Arginine to a Glycine residue in the active site of DCT is the cause of slaty coat colour in mice (Costin *et al.* 2005). Slaty mice have brown/grey eumelanin, caused by a three to four-fold decrease in *DCT* expression (Jackson *et al.* 1992). Research

suggests that both TYRP1 and DCT can act to stabilise TYR, as they form a complex in the membrane of melanosomes, which is thought to be disrupted by TYRP1 and DCT mutations – most notably the brown *TYRP1* mutant (Kobayashi *et al.* 1998; Manga *et al.* 2000; Hirobe *et al.* 2006). The promoters of the TYR gene family are controlled by MITF, which binds to the highly conserved M-box promoter in all three genes (Camacho-Hübner *et al.* 2002; Camp *et al.* 2003).

Another gene which is known to influence the production of melanin is the Membrane Associated Transport Protein (MATP; Newton et al. 2001). Mutations in *MATP* lead to colour dilution, notably the underwhite dilution in mice (Newton *et al.* 2001; Bennett & Lamoreux 2003); buckskin and palomino in horses (Mariat *et al.* 2003) and Oculocutaneous albinism type 4 in humans (Newton *et al.* 2001). It is thought that MATP is involved in TYR transport into the melanosome, because MATP and TYR mutants have similar effects (Rosemblat *et al.* 1994; Newton *et al.* 2001; Graf *et al.* 2007; Tully 2007). This hypothesis is supported by evidence that shows *MATP* to have an expression pattern in embryos similar to that of *TYR* and *TYRP1* (Baxter & Pavan 2002). More recent evidence suggests that MATP may also play a role in pH regulation of the melanosome (Tully 2007). Induced increases in pH have been demonstrated to lead to increased TYR activity (Ancans *et al.* 2001).

Changes in pH have been demonstrated to regulate melanogenesis, influencing the rate at which melanogenesis occurs, eumelanin/pheomelanin ratio and also melanosome maturation (Ancans *et al.* 2001). The melanosome normally has an acidic pH, and neutralisation of this pH promotes eumelanin production (Ancans *et al.* 2001). There is strong evidence that suggests that the P-protein is involved in regulation of pH in melanosomes. Originally it was thought that the P-locus, like MATP, may have been involved in the trafficking of TYR to the melanosomal membrane (Rosemblat *et al.* 1994). However the P-protein shares similarity to the *E. coli* Na+/H+ anti-porter, which suggests that it could play a role in reducing proton concentrations inside the melanosome (Ancans *et al.* 2001). Expression of functional P-protein has been demonstrated to be required for eumelanin production (Rinchik *et al.* 1993).

1.2.3.5 Melanosome Export

The final stage of pigmentation is the transfer of the pigmented melanosomes from the melanocyte, via dendrites, to nearby keratinocytes and hair root cells (Marks & Seabra 2001). Interactions between Ras-related protein Rab27a (RAB27A), Myosin Va (MYO5A) and melanophilin (SLAC2A) will determine the efficiency of the export process. SLAC2A is a linker protein whose n-terminal region (amino-acids 1-146) will bind RAB27A and the middle region (amino acids 241-405) will bind Myo5a (Kuroda *et al.* 2003; Hume *et al.* 2007). The RAB27A/ MYO5A/SLAC2A tripartite complex is necessary for the recruitment of myosin, and the construction of microtubules within dendrites that allows for melanosome transport (Wu *et al.* 2001). Mutations in any of these genes are known to result in dilute phenotypes, such as those observed in *ashen*, *leaden* and *dilute* mice (Wilson *et al.* 2000; Kuroda *et al.* 2003; Hume *et al.* 2007). Some *dilute suppressor* (*DSU*) alleles are known to counteract the effects of the hypomorphic *MYO5A* allele (O'Sullivan *et al.* 2004). It is likely that these *DSU* alleles operate through a phagocytic process, that appears to be independent of MYO5A, or any homologous protein, which allows for more even distribution of melanosomes than observed in MYO5A deficient mice (O'Sullivan *et al.* 2004).

MITF has also been demonstrated to have a role in regulating *RAB27A* expression, binding directly to the *RAB27A* promoter (Chiaverini *et al.* 2008). When MITF is silenced, melanosomes are shown to gather near the nucleus. This is because the protein products of *RAB27A*, *MYO5A* and *SLAC2A* are relocated to the cell body, which renders them ineffective in melanosome export, as opposed to their normal reinforced localisation to the cell periphery (Chiaverini *et al.* 2008).

1.3 Next-Generation Sequencing Technologies

The process through which fibre colour is determined is complex, with 171 genes that affect colour in mice already cloned (http://www.espcr.org/micemut/). The alpaca genome is predicted to be between 2.7 and 3Gb in length (O'Brien *et al.* 2008). Given this complexity, only a broad investigation of all genes that are being expressed in alpaca skin will be efficient to elucidate which genes may be involved

in changing the fibre colour of alpacas. The ability to conduct a study of this magnitude has been made possible through the advent of next-generation sequencing.

1.3.1 Background to Sequencing Technologies

Next-generation sequencing (NGS), also known as massively parallel sequencing, has revolutionised the field of molecular biology. Continuously running 96-well plates on an ABI Prism 3700 series Sanger sequencing platform can generate approximately 900,000bp of sequence per day, however the highest-throughput NGS machine to date, an Illumina HiSeq X can now generate approximately 600 billion bp of sequence per day (www.illumina.com). NGS addressed the limitations of Sanger sequencing by removing the need for gels or polymers as a medium for separating fluorescently labelled DNA fragments, which meant shortening read lengths (Ansorge 2009). Through the introduction and subsequent competition between the various NGS platforms, the cost of sequencing large genomes has become increasingly affordable and available, with even smaller research groups now being able to conduct large-scale genomic work.

1.3.1.1 454 sequencing

The first commercially available NGS instrument was the 454 Genome Sequencer. This was released in 2005 by 454 sequencing technologies (later acquired by Roche) and based on a type of pyrosequencing (Shendure & Ji 2008). In 454 sequencing, emulsion PCRs are utilised, and the amplicons are attached to magnetic beads. These amplicon-bearing beads are then deposited in picolitre-sized wells (which are sized so that only one bead will fit per well); ATP sulfurylase and luciferase are added and when a nucleotide is incorporated light is emitted which is subsequently detected by a sensitive charge coupled device (CCD) camera (Morozova & Marra 2008; Shendure & Ji 2008). Single species of nucleotides are added, one at a time over hundreds of cycles (Morozova & Marra 2008).

1.3.1.2 Solexa sequencing

Other NGS technologies soon followed. In 2006 Solexa (later acquired by Illumina) introduced the sequencing by synthesis idea: bridge PCR where both ends of a library fragment are attached and sequenced on a solid substrate (Ansorge 2009). The

four different nucleotides are labelled with a different florescent dye and depending on which nucleotide is incorporated a different signal is detected by a CCD camera (Ansorge 2009).

1.3.1.3 SOLiD Sequencing

The SOLiD system from ABI was released in 2007. It also utilises emulsion PCR, which is bound to beads before being deposited on a dense, disordered array (Ansorge 2009). Octamer oligonucleotides are added and cleaved at specific positions with florescent markers at the end of the oligonucleotide being detected, determining the base position of the two nucleotides at the end of the octomer (Morozova & Marra 2008).

SOLiD sequencing has now been superseded by Illumina's short-read technology. A large part of this is likely due to the increased read lengths given by Illumina platforms, which allow for improved assembly. Another large part is likely to be due to the "closed system" where SOLiD reads are output in colourspace which makes them mostly incompatible with many open-source software programs commonly used for analysis. With the advent of the SOLiD 5500xl platform the output was in xsq format (Zhang *et al.* 2011), which was only compatible with the LifeScope proprietary software, reducing researcher flexibility to analyse data by alternative software.

1.3.1.4 Illumina Sequencing

There have been a number of new iterations of the Illumina technologies leading to the modern HiSeq instruments which utilise paired end 150bp reads and can produce up to 1500 Gb of data per run on a single machine. The X10 series features 10 such linked HiSeq instruments, which can produce up to 1.8 Tb of data per run. These modern Illumina technologies still utilise similar sequencing-by-synthesis technologies to the first Illumina sequencers (www.illumina.com).

1.3.1.5 Third-generation sequencing technologies

More recent trends have moved toward single molecule sequencing, also known as third-generation sequencing technologies. These have been headlined by the PacBio

sequencing technology, which do not generate as much data as many of the second generation sequencing technologies, but make up for this with vastly improved read lengths, which can extend up to 60Kb in a single read (Rhoads & Au 2015). PacBio RS instruments achieve this extended read length through single-molecule real-time (SMRT) sequencing. SMRT sequencing utilises zero-mode wavelength technology, in which a single molecule of polymerase is fixed to a solid substrate and is observed as it incorporates fluorescently labelled dNTPs into a fragment which has been circularised in library preparation (Rhoads & Au 2015). Individual cycles result in large numbers of sequencing errors, however the polymerase is observed over the course of many rounds of replication and the consensus sequence is taken, minimalizing these errors (Rhoads & Au 2015). An overview of the current capacities of sequencing technologies can be found in Table 1.5

Table 1.5. Comparison of the key features of major next generation sequencing platforms. Adapted from Shokralla *et al.* (2012) and Rhoads and Au (2015)

	Roche 454 GS FLX	**Illumina GAIIx**	**AB SOLiD 5500xl**	**Illumina HiSeq 2000**	**Illumina HiSeq X**	**PacBio RS**
Read lengths (bp)	400-500	50-75	35-75	100-200	150bp	>60kb (average ~10kb)
Number of reads per run	1×10^6	6.4×10^8	6.0×10^9	6.0×10^9	6.0×10^9	50,000
Sequencing output per run	500Mb	95Gb	250Gb	600Gb	1.6-1.8Tb	500Mb
Run time	10 hours	7.5-14.5 days	7-8 days	11 days	3 days	3 hours
Advantages	Longer reads	High quality reads, high output	High quality reads, high output	High quality reads, very high read numbers and output	High quality reads, very high read numbers and output	Longest reads
Disadvantages	Low output	Short reads	Short reads	Short reads – compared to PacBio	Short reads – compared to PacBio	Relatively low numbers of reads/ output, Poor quality reads require error correction

1.3.2 Sequencing RNA

The advent of next-generation sequencing has allowed the sequencing of a whole transcriptome in one experiment, known as RNAseq (Wang *et al.* 2009). RNAseq involves sequencing all transcripts that are expressed within a sample, allowing the quantitation of different transcripts, identification of novel transcripts, and prediction of the structure of all mRNA present at a given time in a sample (Ansorge 2009; Wang *et al.* 2009). RNAseq is conducted by taking a population of RNA, and reverse transcribing it to create a cDNA library (Ansorge 2009; Hrdlickova *et al.* 2016). Adapters are then added before sequencing using the appropriate procedure on an NGS platform (Wang *et al.* 2009). As RNA transcripts will be expressed at different levels inside the cell this will in turn result in different proportions of the respective cDNAs being created from the RNA transcripts present in the library. These will then theoretically correspond to the amount of reads generated for the given cDNA and thus its expression can be determined (Wang *et al.* 2009; Hrdlickova *et al.* 2016).

The type of library preparation for RNAseq can either be stranded or unstranded, depending on requirements. Stranded RNAseq causes libraries that are generated to be sequenced in the direction of transcription on the chromosome strand (McGettigan 2013). The advantage of stranded libraries is the ability to determine whether a transcript runs in a particular orientation, which is necessary to obtain accurate data when measuring gene expression levels where there are overlapping transcripts on both strands of a chromosome, or where there is anti-sense transcription.

The type of RNA that is used can be total or fractionated. Commonly the type of RNA used in RNAseq is fractionated, as up to 85% of mammalian total RNA is rRNA, which has limited use in transcriptome studies (Morlan *et al.* 2012). Methods such as ribo-depletion remove the majority of rRNA from the total RNA so that the RNA of interest, in particular mRNA, is represented at a higher proportion (Wang *et al.* 2009; Morlan *et al.* 2012). Poly-A binding is a commonly used method when only mRNA is required and methods aim to remove all non mRNA from the total RNA pool (Wang *et al.* 2009). One advantage of ribo-depletion over a method such as poly-A binding is that all the small RNAs such as miRNAs remain in the sample.

Prior to the advent of next-generation sequencing, microarray was the predominant way of analysing mammalian transcriptomes. Microarray is based on the hybridisation of genomic sequences to a pre-defined set of probes which exist on a chip (Zhao *et al.* 2014). It relies heavily on the target sequences to use as probes, whereas RNAseq has no or very little reliance on genomic reference and annotation sequences (though good annotations assist the analysis) and subsequently is able to detect novel transcripts (Zhao *et al.* 2014). RNAseq has single-base resolution, allowing for detection of SNPs and different isoforms, many of which could not be detected by microarray (Wang *et al.* 2009). Other advantages of RNAseq over microarray include the absence of background noise caused by cross-hybridisation of probes, higher sensitivity, particularly for the quantification of genes expressed at very low or very high levels; and higher accuracy in determining gene expression differences (Wang *et al.* 2009; Zhao *et al.* 2014; Table 1.6).

Table 1.6 Comparison of microarray vs RNAseq (adapted from Wang et al. 2009)

	Microarray	**RNAseq**
Principle	Hybridisation	Sequencing
Resolution	Several to 100bp	Single base
Reliance on Genomic Sequence	Yes	Limited, not a requirement
Background Noise	High	Low
Ability to Distinguish Different Alleles/Isoforms	Dependent on the location of the probes and amount of variation.	Yes

An alternate way to look, at a global scale, at genes that cause a change to phenotype is through sequencing the exome or via linkage studies such as genome-wide association studies (GWAS). Exome studies are also able to determine the coding sequence of genes, however there is no exome kit applicable to alpacas, and although the capabilities to find SNPs in coding regions of genes would be vastly increased due to the evenness of coverage, no information regarding gene expression would be generated. Further, the coding sequence of a number of key alpaca genes have already been determined across different coloured animals in candidate gene studies

(Powell *et al.* 2008; Feeley & Munyard 2009; Cransberg & Munyard 2011; Feeley *et al.* 2011). GWAS studies are another way to look for complex traits in a population. The principle of GWAS is based on linkage disequilibrium, where alleles are non-randomly associated with other alleles, forming haplotypes (Visscher *et al.* 2012). Certain SNPs can be good markers, or tags, for particular haplotypes, even when the SNP lies well outside the causative gene for a particular trait (though still relatively close in a genomic context; Visscher *et al.* 2012). GWAS represents a valid alternative way of studying the genetics of pigmentation in alpacas.

1.3.3 Reference Genomes

In 2008 the first reference genome for alpacas, named vicPac1, was released (GCA_000164845.1). It was released as part of the mammalian genome project, which provided low coverage Sanger assemblies of 24 mammals. To improve this assembly, 454 sequencing data was generated from the same animal, and included fragment and 3kb paired-end sequences, which was then assembled using Newbler v2.7 (454 Life Sciences) to create vicPac2 (GCA_000164845.3; Table 1.7). Annotations for vicPac2, based on Gnomon gene predictions (Souvorov *et al.* 2010) were released in late 2015.

Table 1.7 Assembly metrics and comparison of the vicPac1 and vicPac2 genomes.

	vicPac1	**vicPac2**
Size	1.92gb	2.17gb
Scaffold N50	230kb	7.2mb
Contig N50	3.9kb	24kb
Predicted Genes	15236	24540
Predicted Coverage	2.5×	22×
Sequencing Technologies Used	Sanger	Sanger And 454

1.4 Aims and Purpose

Melanogenesis is regulated by the interactions between hundreds of colour genes. This complex network of interactions causes the extensive variety of colours seen in mammals. In alpacas we are only beginning to elucidate the regulation of colour and this project will expand the base of knowledge in this area. We have utilised next-generation sequencing technologies to discover differences in expression of genes in alpacas which are known to be involved in the determination of fibre colour in other species, as well as other parts of the alpaca skin transcriptome which may vary in different coloured alpacas. The general hypothesis is that changes in the alpaca skin transcriptome will result in a change of fibre colour.

Overall this project aims to:

- Characterise the melanins present in alpaca fibre.
- Determine whether RNAseq is a useful tool for analysing colour genes in the alpaca skin transcriptome.
- Investigate the alpaca skin transcriptome.

The purpose of this project is to bring new knowledge about colour genetics to an industry where colour is a commercially important trait. More knowledge will allow breeders to better breed for select traits, which will help smaller breeders in particular - which make up the majority of Australian breeders, to meet quotas of uniformly coloured fibre

2

Melanin Characterisation in Australian Alpacas

"If you don't know where you are going, you might wind up someplace else."

Yogi Berra

Please note that this chapter contains material from the following published paper: Cransberg, Rhys, Kazu Wakamatsu, and Kylie Munyard. "Melanin characterisation suggests that the brown phenotype in alpaca (Vicugna pacos) is predominantly pheomelanic." *Small Ruminant Research* 114.2 (2013): 240-246.

2.1 Introduction

Mammalian colour is caused by black to brown eumelanin and red to yellow pheomelanin (Wright 1917; Ito & Wakamatsu 2003). The combination and distribution of melanins in keratinocytes and hair determines colour (Oyehaug *et al.* 2002; Hoekstra 2006). In mammals, red to brown coat colours may be caused by pheomelanin such as seen in Irish setter dogs (Newton *et al.* 2000) and chestnut horses (Sponenberg *et al.* 1988) a combination of pheomelanin and eumelanin such as in brown Asiatic sheep (Aliev *et al.* 1990), or a modified form of eumelanin, which results in a distinctive colour, such as is seen in brown mice and chocolate Labradors (Zdarsky *et al.* 1990; Schmutz *et al.* 2002).

This study aims to characterise the melanin content in alpaca fibre. Knowing the melanin types and content of alpaca fibre will allow better predictions of the genes that may be responsible for various alpaca fibre colours, as well as providing an objective method for assigning phenotype in alpacas.

An estimate of total melanin and a ratio of eumelanin to pheomelanin in fibre can be obtained using simple spectrophotometric methods (Ozeki *et al.* 1996). In addition, high performance liquid chromatography (HPLC) can be used to determine whether there are significantly higher levels of eumelanin in any mixed melanin samples, and if the eumelanin present is DHI or DHICA derived (Wakamatsu & Ito 2002). The alkaline peroxide oxidation of eumelanins produces pyrole-2,3,5-tricarboxylic acid, (PTCA) and to a lesser extent pyrole-2,3-dicarboxylic acid (PDCA; Ito & Wakamatsu 2003). PTCA is derived primarily from DHICA units whereas PDCA indicates DHI eumelanin units (Ito & Wakamatsu 1998). Pheomelanic markers that result from this oxidation are 4-amino-3-hydroxyphenylalanine (4-AHP) and its isomer 3-amino-4-hydroxyphenylalanine (3-AHP) as well as thiazole-2,4,5-tricarboxylic acid (TTCA).

In this study a combination of spectrophotometric and HPLC analysis has been utilized to determine the relative amount of eumelanin and pheomelanin in various colours of alpaca fibre. The investigation will provide information about the pigments that are being produced in different alpaca colour phenotypes, and therefore the possible genetic cause of these phenotypes.

2.2 Materials and Methods

2.2.1 Samples

The spectrophotometric analysis of melanins included 45 alpaca fibre samples, five of each colour (Table 2.1), which were donated by Australian Alpaca Fleece Ltd. (AAFL), and 86 samples which were collected from the body under the left elbow of animals exhibited at the Canning Vale Winter Alpaca Show 2009 (Western Australia). Samples were assigned to one of 11 colours based on AAFL classification, or the show class entered (Table 2.1). Samples were collected from all animals at the show where permission to sample was granted (i.e. there was no selection). Thirteen dog hair samples from known eumelanic brown genotypes included English Springer-Spaniel (5), Chocolate Labrador (2), Chocolate Poodle (4), and Liver-Spotted Dalmatian (2).

Table 2.1 Number of samples tested for each colour phenotype and the amount of alpaca fibre used for spectrophotometric analysis.

Fibre Colour	Number of Animals	Amount Fibre (mg)
White	22	20mg
Light Fawn	21	20mg
Fawn	23	15mg
Light Brown	8	10mg
Brown	10	10mg
Dark Brown	7	5mg
Red/Brown	5	5mg
Black Brown	7	3mg
Black	28	2mg

The HPLC analysis was conducted on 75 alpaca samples. These consisted of the 45 AAFL alpaca fibre samples, which were also analysed using spectrophotometric methods, (five each of white, light fawn, fawn, light brown, brown, dark brown, red brown, black brown, and black) as well as five each from AAFL of the rosegrey and grey phenotypes. Another twenty samples were collected from animals on farms in Western Australia, consisting of 5 samples from each of pink-skin fawn, dark-skin fawn, chestnut and silvergrey phenotypes.

For the analysis of the uniformity of colour over the whole body, samples were collected from six body locations on each of five white, brown and black animals: the neck, shoulder, body immediately under the elbow, mid-side, top-saddle and thigh of the left side of the animals. Animals were from farms in Western Australia and all had a phenotype that is usually described as solid (i.e. unpatterned).

2.2.2 Spectrophotometric Analysis

Fibre samples were washed and ground into a fine powder in a mortar and pestle under liquid nitrogen. Both alpaca and dog samples (described above) were analysed using the same technique. Triplicate aliquots were prepared and assessed using the method of Ozeki *et al.* (1996). Samples were centrifuged and the absorbance of the supernatant determined at 500nm and 650nm using a Shimadzu UVmini-1240 spectrophotometer. Different amounts of fibre were added, based on their colour, to achieve spectrophotometric absorption in the optimal range for the instrument (Table 2.1).

To assess the effect of potential contaminants on the hair samples, each of the common possible contaminants was tested on a single white hair sample. The contaminants tested were observed in some samples before washing, or were present on farms, these included black fibre (two strands), grass, dried grass, black dirt, red dirt and grey dirt. The fibre was exposed to the potential environmental contaminants by either rubbing a portion of fibre vigorously against the contaminant in the case of dirt, or the addition of a small amount to the pre-ground sample, in the case of fibre and grass.

2.2.3 HPLC Analysis

HPLC analysis was used to quantify the pyrrole-2,3,5-tricarboxylic acid (PTCA), pyrrole-2,3-dicarboxylic acid (PDCA), thiazole-2,4-5-tricarboxylic acid (TTCA), 4-amino-3-hydroxyphenylalanine (4-AHP) and 3-amino-4-hydroxylphenylalanine (3-AHP) content of the fibre samples. The eumelanin markers PTCA and PDCA were prepared as described in Ito and Wakamatsu (1998) and TTCA was prepared as

described in Wakamatsu *et al.* (2003) with minor modifications to improve the yields. The preparation of 4-AHP was completed as described in Wakamatsu and Ito (2002). 3-AHP was purchased from Sigma-Aldrich (St. Louise, MO. USA). Alkaline H_2O_2 oxidation to measure eumelanin (as PTCA and PDCA) and pheomelanin (as TTCA) was performed as described in Wakamatsu *et al.* (2009). In brief, 100 µl of water suspensions (1.0 mg fibre) was placed in 10-ml screw-capped conical rest tubes, to which 375 µl 1 mol/l K_2CO_3 and 25 µl 30% H_2O_2 (final concentration: 1.5%) were added. The test tubes were mixed vigorously at 25 ± 1°C for 20h on a test-tube mixer. The residual H_2O_2 was decomposed by adding 50µl 10% Na_2SO_3 and the mixture was then acidified with 140 µl 6 mol/l HCl. Each reaction mixture was centrifuged at 4000 g for 1 min, and an aliquot (80 µl) of each supernatant was directly injected into the HPLC system. HPLC oxidation products were analysed with an HPLC system consisting of a JASCO 880-PU liquid chromatograph (JASCO Co., Tokyo, Japan), a Shiseido C18 column (Capcell Pak MG, 4.6 x 250 mm, 5 µm particle size; Shiseido, Tokyo, Japan), and a JASCO UV detector at 269 nm (JASCO Co., Tokyo, Japan). The mobile phase was 0.1 mol/l potassium phosphate buffer (pH 2.1)/methanol, 99:1 (v/v). HI reductive hydrolysis to measure pheomelanin (as 4-AHP and 3-AHP) was performed as described in Wakamatsu *et al.* (2002).

2.2.4 Statistical Analysis

Different colours of fibre were compared using the Students *t*-test, statistical significance was accepted at $p \leq 0.05$.

2.3 Results

2.3.1 Fibre Uniformity Analysis

Differences were seen in the A_{650}/A_{500} ratio of the fibre samples taken from different locations in white animals (P=0.002; Figure 2.1). The largest difference was between the neck samples and shoulder samples (P=0.013) and samples taken from behind the elbow (P=0.002). There was no significant difference in the A_{650}/A_{500} ratio of samples collected from different locations on brown (P=1.000; Figure 2.2) or black (P=0.772; Figure 2.3) animals.

There was no significant difference in the total melanin (i.e. A_{500}/mg values) between all positions taken from white animals (P=0.569; Figure 2.4). There were, however, significant differences between samples taken from different positions in brown (P=0.022; Figure 2.5) and black animals (P=0.036; Figure 2.6). In both these colour groups the sample taken from behind the elbow had a significantly lower total melanin concentration compared to at least one other sample location.

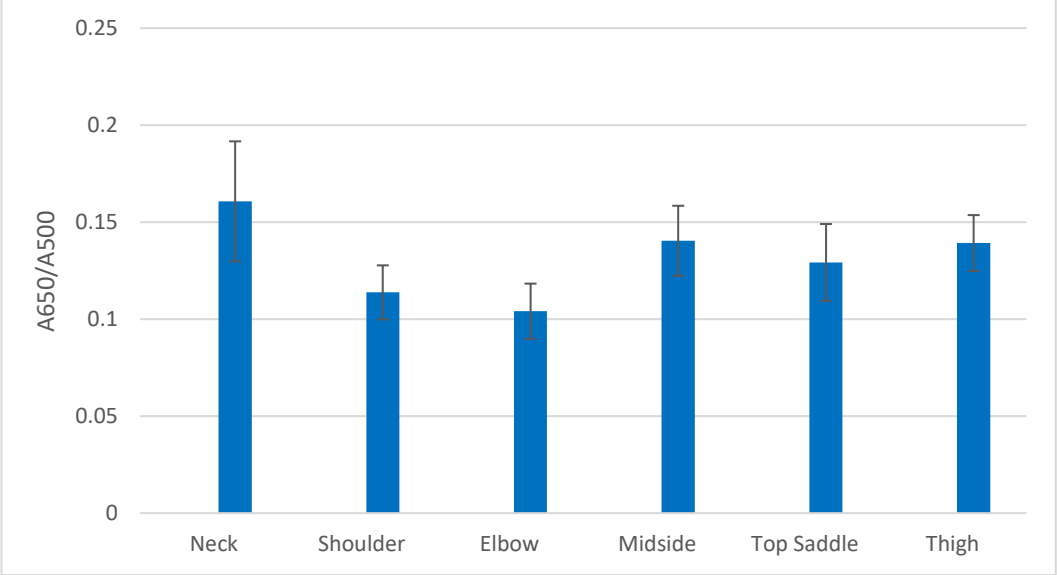

Figure 2.1 Uniformity of melanin composition in white animals. Pheomelanin vs. eumelanin ratio in alpaca fibre sampled from different positions on the animal. Each position value is the average ratio of absorbance at 650nm vs. the absorbance at 500nm from five samples. Values ≤ 0.15 indicate predominantly pheomelanic fibre; values ≥ 0.25 indicate predominantly eumelanic fibre and values in between these ratios indicate that neither eumelanin nor pheomelanin is predominant. Error bars represent 1 standard deviation from the mean.

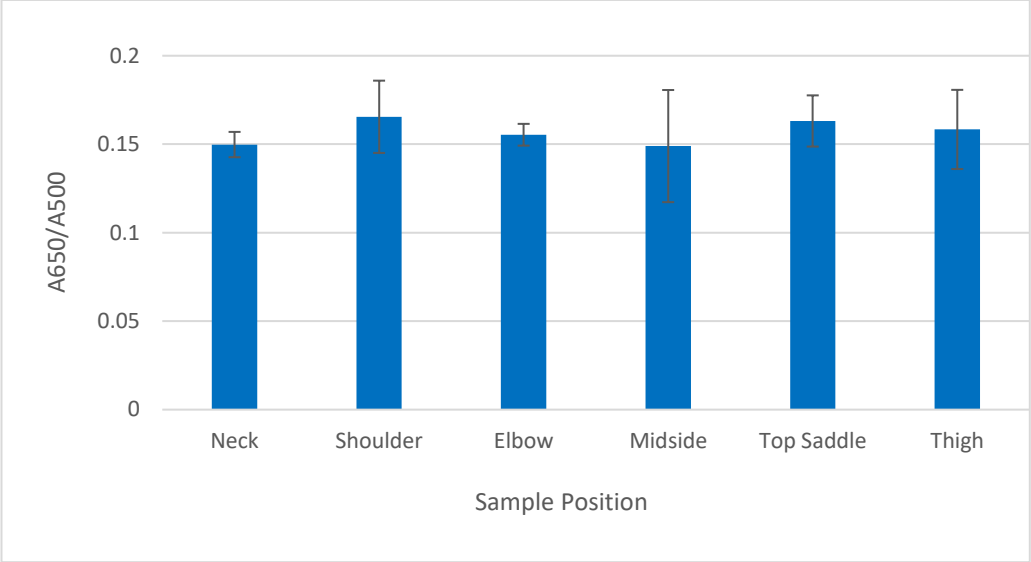

Figure 2.2 Uniformity of melanin composition in brown animals. Pheomelanin vs. eumelanin ratio in alpaca fibre sampled from different positions on the animal. Each position value is the average ratio of absorbance at 650nm vs. the absorbance at 500nm from five samples. Values ≤ 0.15 indicate predominantly pheomelanic fibre; values ≥ 0.25 indicate predominantly eumelanic fibre and values in between these ratios indicate that neither eumelanin nor pheomelanin is predominant. Error bars represent 1 standard deviation from the mean.

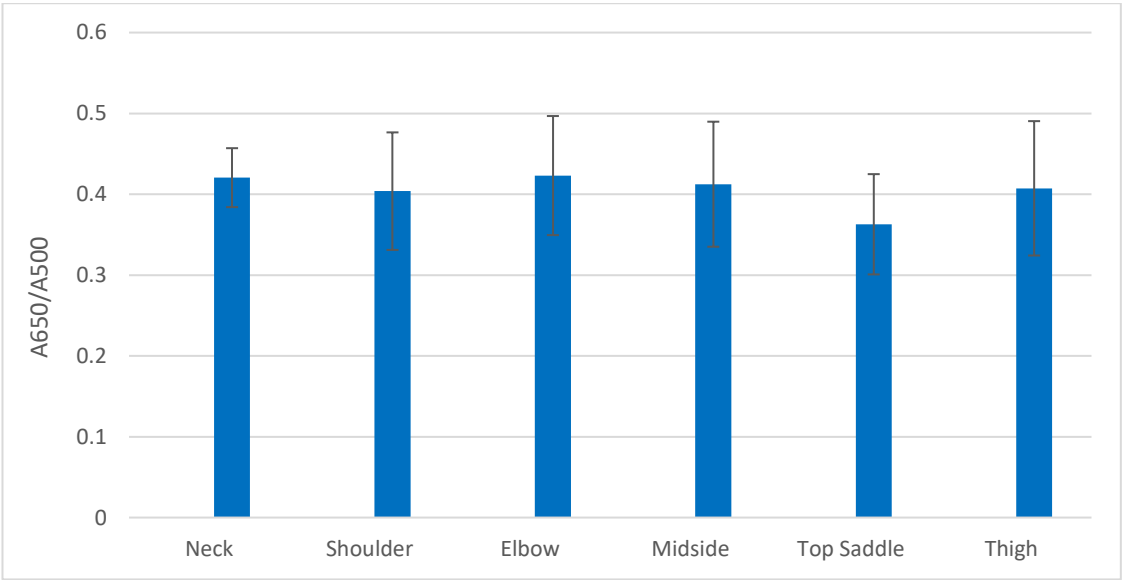

Figure 2.3 Uniformity of melanin composition in black animals. Pheomelanin vs. eumelanin ratio in alpaca fibre sampled from different positions on the animal. Each position value is the average ratio of absorbance at 650nm vs. the absorbance at 500nm from five samples. Values ≤ 0.15 indicate predominantly pheomelanic fibre; values ≥ 0.25 indicate predominantly eumelanic fibre and values in between these ratios indicate that neither eumelanin nor pheomelanin is predominant. Error bars represent 1 standard deviation from the mean.

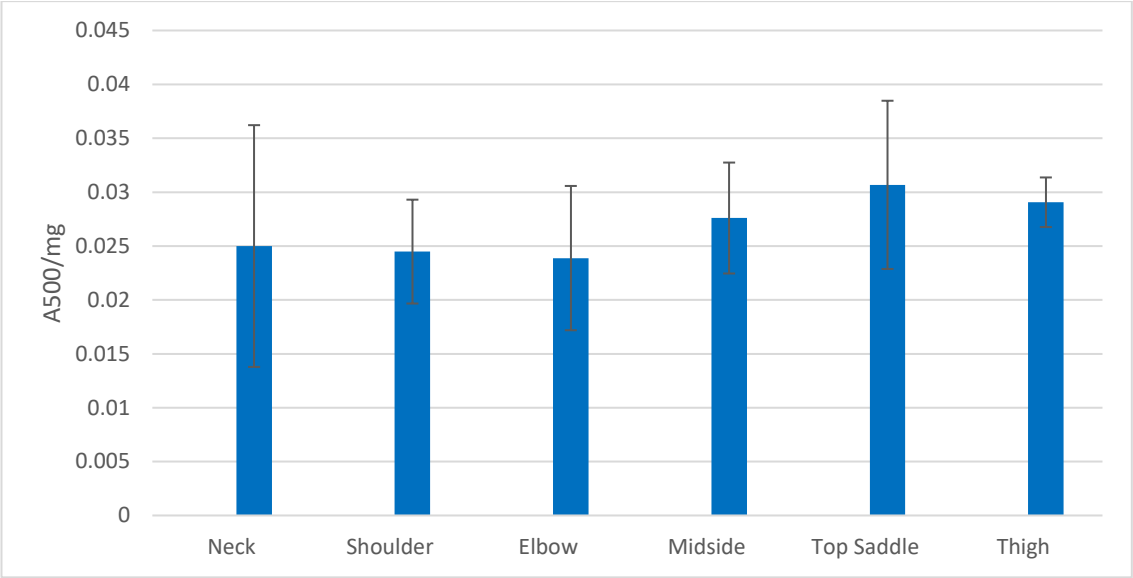

Figure 2.4 Uniformity of total melanin content in white animals. Total melanin concentration of alpaca fibre taken from different positions on the animal, determined by A_{500} values per mg of fibre used. Each position value is an average of five samples. Error bars represent 1 standard deviation from the mean.

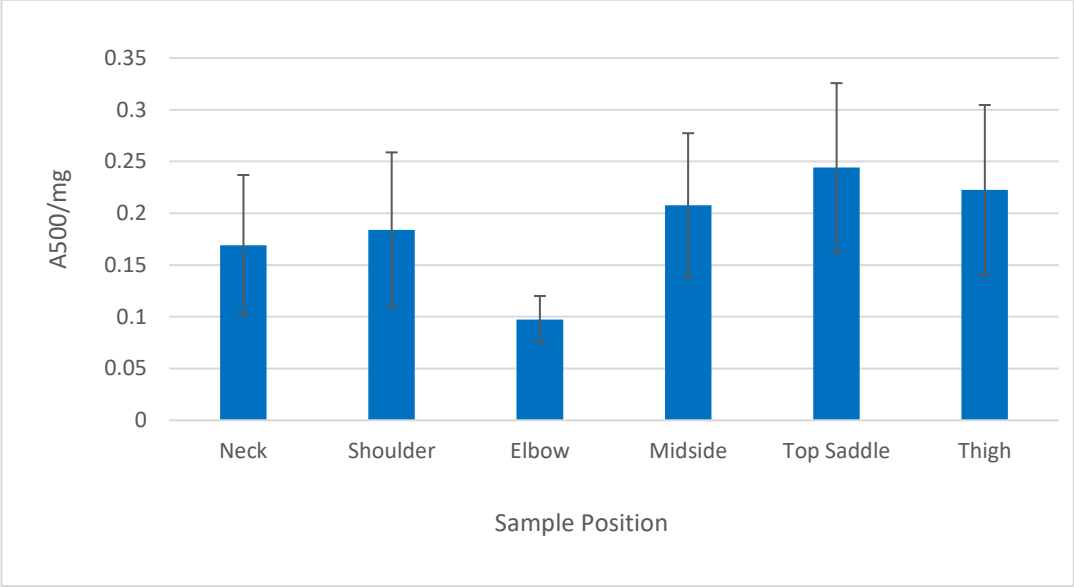

Figure 2.5 Uniformity of total melanin content in brown animals. Total melanin concentration of alpaca fibre taken from different positions on the animal, determined by A_{500} values per mg of fibre used. Each position value is an average of five samples. Error bars represent 1 standard deviation from the mean.

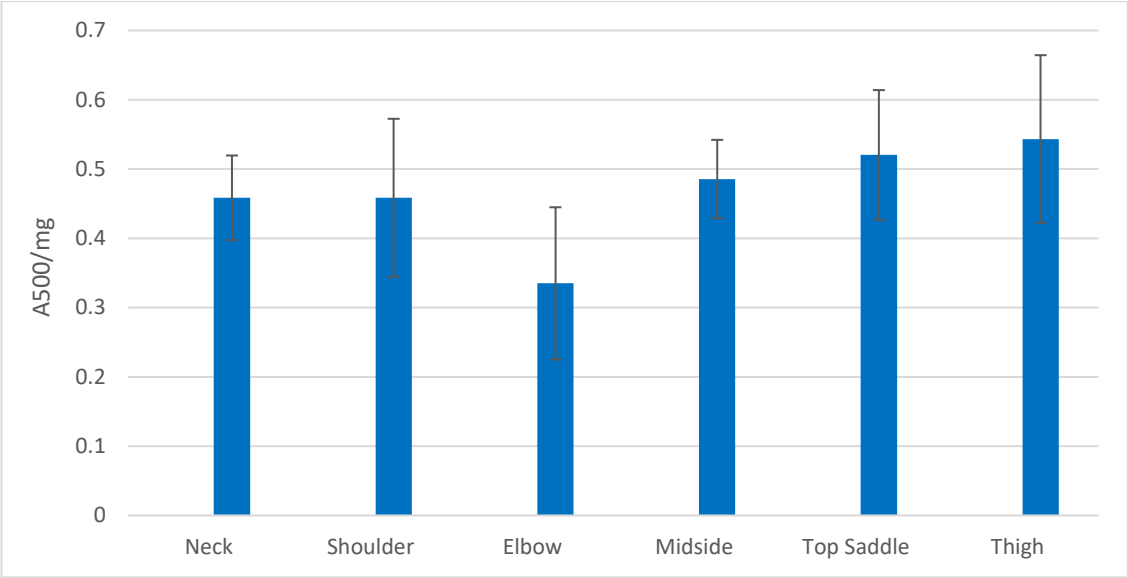

Figure 2.6 Uniformity of total melanin content in black animals. Total melanin concentration of alpaca fibre taken from different positions on the animal, determined by A_{500} values per mg of fibre used. Each position value is an average of five samples. Error bars represent 1 standard deviation from the mean.

2.3.2 Effect of Contaminants

The effect that various natural contaminants may have on the A_{650}/A_{500} ratio and the A_{500} value was measured (Table 2.2). All contaminants had a large effect on the values recorded in white fibre.

Table 2.2 The effect that potential contaminants have on the measurement of the amount and ratio of melanins. Each contaminant was tested on a single sample, which was exposed to the potential environmental contaminants by either rubbing a portion of fibre vigorously against the contaminant in the case of dirt, or the addition of a small amount, in the case of fibre and grass.

Hair Color	Contaminant	A_{650}/A_{500} Ratio	A_{500}/Mg
White	None	0.090	0.049
White	Black Fibre (two strands)	0.195	0.083
White	Grass	0.734	0.139
White	Dried Grass	0.619	0.155
White	Black Dirt	0.698	0.171
White	Red Dirt	0.163	0.101
White	Grey Dirt	0.560	0.144

2.3.3 Spectrophotometric Fibre Analysis

Spectrophotometric analysis was completed on the 131 samples collected from AAFL and the Canning Vale Winter Alpaca Show 2009. There was no significant difference in the total amount of melanin between white and light fawn fibre (P=0.209). However white had significantly lower amounts of total melanin when compared to fawn (P=0.030), although total melanin content of light fawn was not significantly different from that of fawn (P=0.148; Figure 2.7). With the increase in intensity of colour, for example dark brown compared to light brown; the total amount of melanin observed also increased (Figure 2.7).

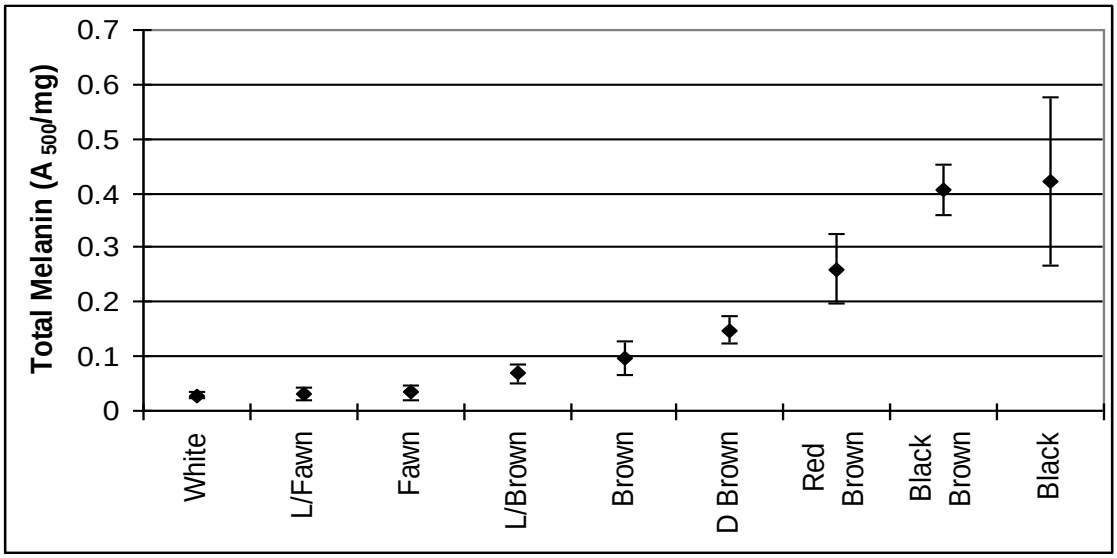

Figure 2.7 Total melanin concentration of different colours of alpaca fibre, determined by A_{500} values per mg of fibre used. Values represent an average of three replicates of each of five samples per colour. Error bars represent 1 standard deviation from the mean.

The A_{650}/A_{500} ratio demonstrates that the relative proportion of eumelanin to pheomelanin in white fibre was not significantly different from light fawn (P=0.238) in the types of melanin present, nor was it different from fawn (P=0.428). Similarly, light fawn and fawn were not significantly different from each other (P=0.201; Figure 2.8). Interestingly, when white was compared to AAFL brown phenotypes, no significant difference was observed between white, brown (P=0.208) and red brown (P=0.232), indicating that the major difference between the whites and these browns was the amount of pigment, as opposed to a change in the relative levels of pheomelanin and eumelanin being produced. Compared to the other brown

phenotypes, on average dark brown alpaca phenotypes had a greater relative pheomelanin concentration (A_{650}/A_{500} = 0.097) and light brown phenotypes a greater relative eumelanin concentration (A_{650}/A_{500} = 0.187; Figure 2.8).

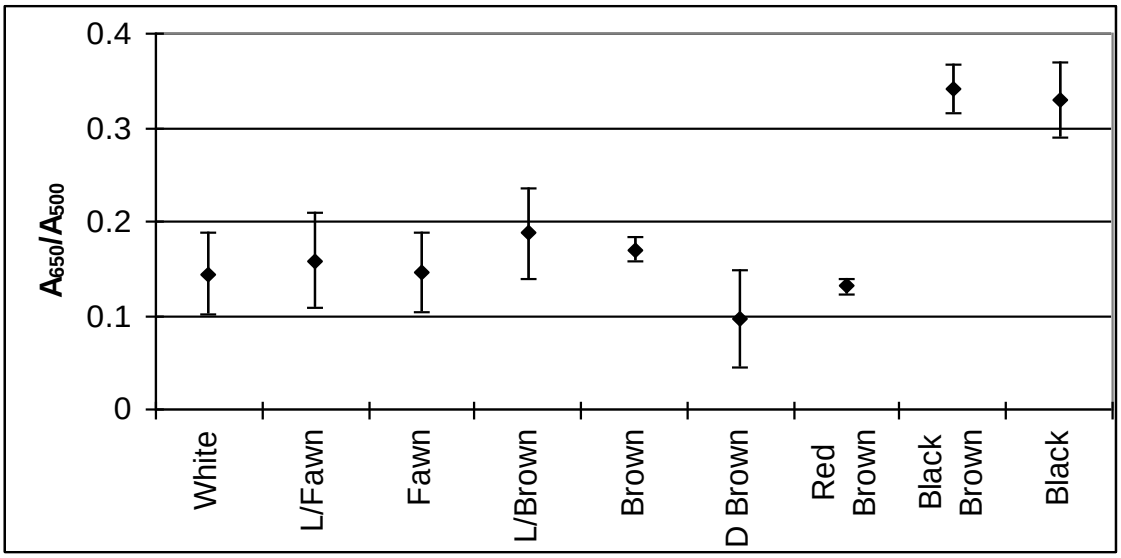

Figure 2.8 Pheomelanin vs. eumelanin ratio in alpaca fibre of different colours. Values represent the average ratio of absorbance at 650nm vs. the absorbance at 500nm for three replicates of each of five samples per colour. Error bars represent 1 standard deviation from the mean. Values ≤ 0.15 indicate predominantly pheomelanic fibre; values ≥ 0.25 indicate predominantly eumelanic fibre and values in between these ratios indicate that neither eumelanin nor pheomelanin is predominant.

When the brown alpaca fibre samples were compared to hair from a known eumelanic brown phenotype, in this case several breeds of dog, it was evident that none of the alpaca samples were predominantly eumelanic (Figure 2.9). Brown alpacas clearly have a mixed or predominantly pheomelanic melanin content. All eumelanic dog samples gave A_{650}/A_{500} ratios greater than 0.277.

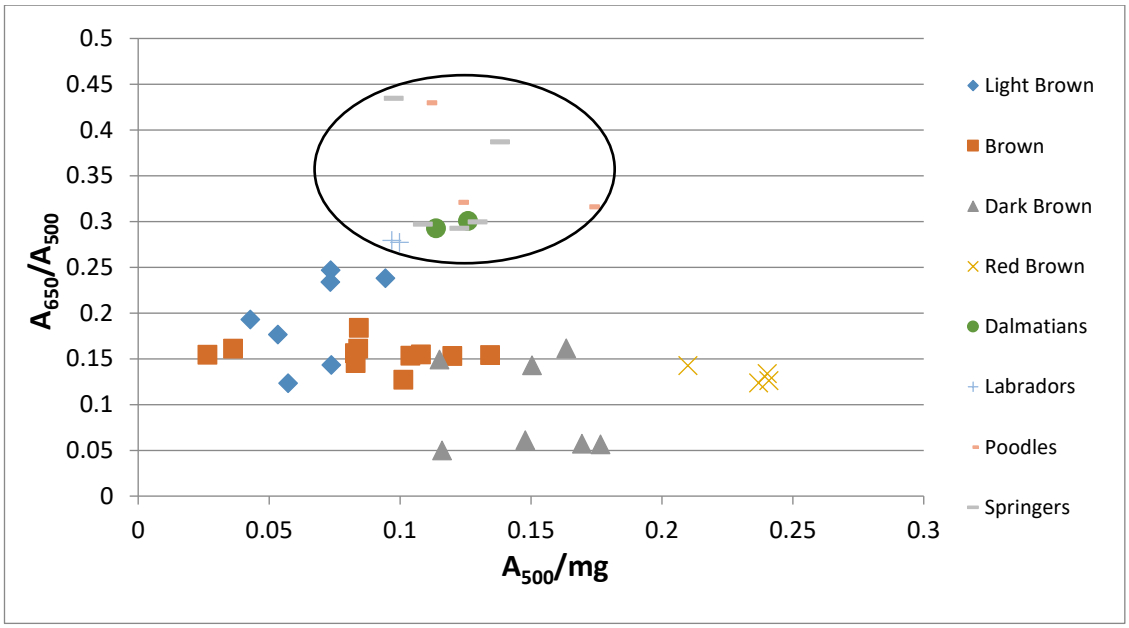

Figure 2.9 Comparison of the total amount of melanin (A_{500}/mg) and relative proportion of eumelanin and pheomelanin ($A_{650/500}$) between brown alpacas and known eumelanic brown dogs. Values represent average A_{500} values per mg of fibre for three replicates per sample plotted against average A_{650}/A_{500} ratio for three replicates of the same sample.

2.3.4 HPLC Fibre analysis

HPLC analysis supported the results of the spectrophotometric analysis. The black, black brown, grey and silvergrey animals were all predominantly eumelanic, while red brown, dark brown, chestnut, brown and light brown were all predominantly pheomelanic (Figure 2.10). Rosegrey had the most even proportions of both eumelanin and pheomelanin (Figure 2.10). Most of the variation between groups occurred in the intensity of the predominant type of melanin, as opposed to a change in melanin type (Figure 2.10). Both pheomelanic and eumelanic animals had only low amounts of PDCA eumelanin which had a moderate positive correlation to PTCA eumelanin levels (Figure 2.11).

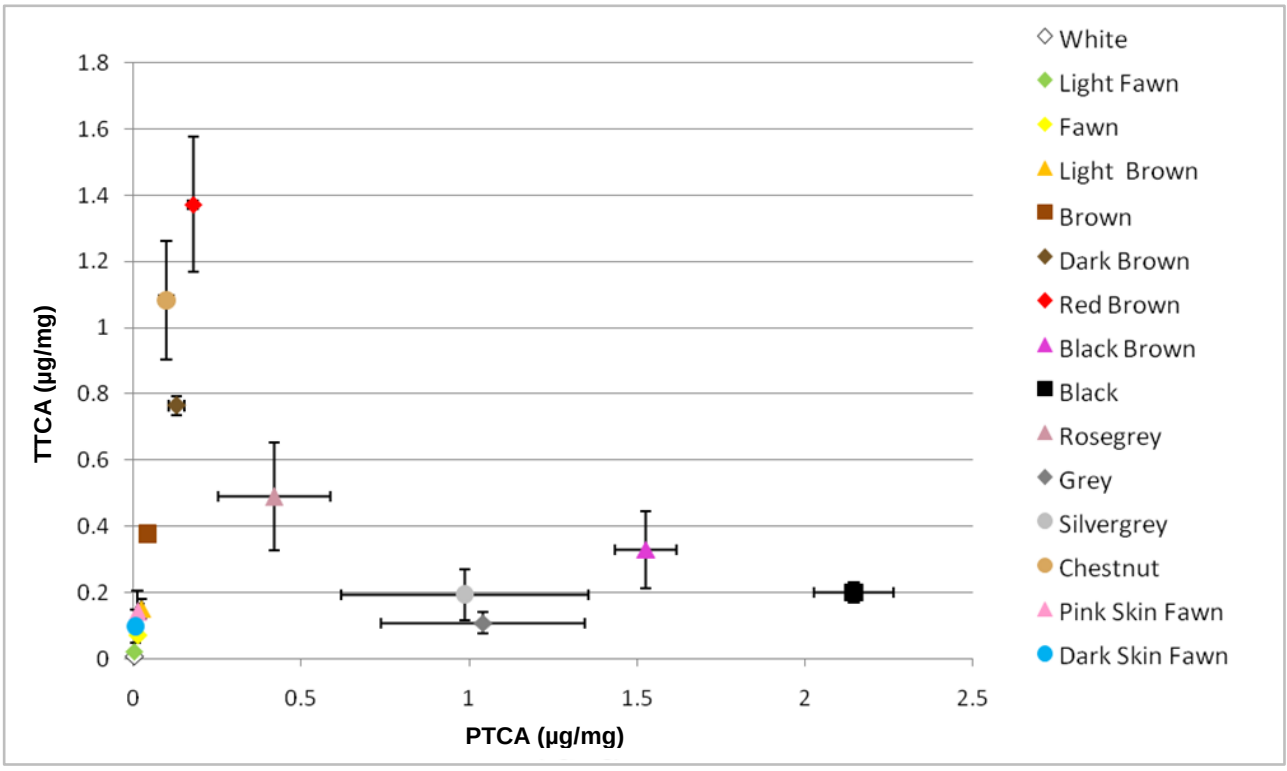

Figure 2.10 Comparison of TTCA and PTCA content. TTCA is an indicator of pheomelanin and PTCA is an indicator of DHICA-derived eumelanin. Values are representative of the average of five different animals. Error bars represent one standard error from the mean.

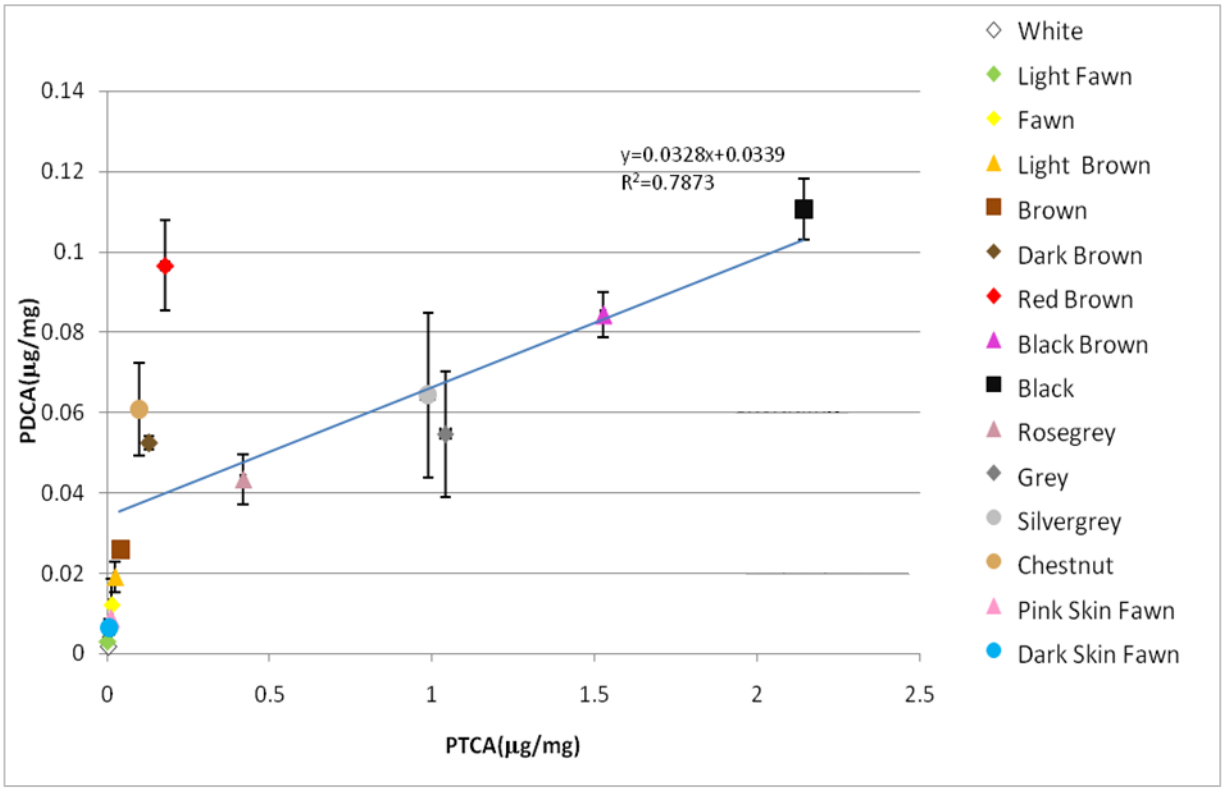

Figure 2.11 Comparison of PTCA (an indicator of DHICA-derived eumelanin) to PDCA (an indicator of DHI-derived eumelanin). Values are representative of the average of five different animals. Error bars represent one standard error from the mean.

2.4 Discussion

These data demonstrate that the fibre of all of the brown alpacas analysed in this study was predominantly pheomelanic, or contained a mixture of pheomelanin and eumelanin, and was not brown due to eumelanic brown pigment. Therefore, as opposed to a phenotype which results from brown eumelanin production, most likely associated with a mutation in the *TYRP1* gene, it is likely that in most cases brown in alpacas is caused by a genotype consisting of a wild-type melanocortin-1 receptor (*MC1R)* gene, coupled with a variety of agouti signalling protein (*ASIP*) variant alleles. The result of this is either a pheomelanic animal with some eumelanin production around the extremities, as seen in bay horses (Thiruvenkadan *et al.* 2008), or by a non-functional MC1R which results in pheomelanin production, as seen in chestnut horses (Marklund *et al.* 1996).

Alpaca breeding records include multiple instances where two brown parents produce black offspring (Paul 2002), which is theoretically impossible if the recessive *TYRP1* brown mutation is acting to cause the brown phenotype, because eumelanic brown is a recessive genotype. It has also been reported that two black parents can produce brown offspring (Paul, 2002), suggesting it is possible that the *TYRP1* brown mutation does exist in alpacas. However recent progress in alpaca colour genetics has demonstrated that two black parents can produce brown cria when the two parents have non-functional *ASIP* alleles, whilst being heterozygous for a non-functional *MC1R* allele (Feeley et al., 2011). This is contrary to the view that black is dominant over brown in alpacas (Valbonesi *et al.* 2011). The assumption that mating two black animals to produce a brown animal is proof that brown is recessive in this species is incorrect when you consider a possible non-functional genotype at *MC1R*, or potential misidentification of extreme *ASIP* $a^t a$ genotype parents (a very dark black-brown phenotype; Munyard 2011) as an aa genotype (black phenotype).

These data show that, an alteration in the amount of pheomelanin is responsible for the difference in colour between white and brown animals, with brown animals having significantly more pheomelanin than white animals (Figure 2.7). This is in direct contrast to results from a previous study, which concluded that pheomelanin

content had no relationship to alpaca fibre colour (Fan *et al.* 2010)**.** However the authors noted that the amount of pheomelanin that they detected was very low – which may have prevented the detection of any differences that may have been present.

Some brown samples gave mixed values (0.15-0.25) for the A_{650}/A_{500} ratio (Figure 2.9). It is likely that an *MC1R*/*ASIP* allele combination is acting to produce brown fibre caused by a mixture of eumelanin and pheomelanin, as previously demonstrated in brown sheep (Aliev *et al.* 1990). Mixed melanogenesis is a well-established concept in pigmentation (Ito and Wakamatsu, 2008). Light brown alpacas, in particular, were observed to have A_{650}/A_{500} ratios that indicated mixed melanin content. Mixed melanogenesis may be observed because of the presence of a few eumelanic hairs in these animals. It is possible that similar levels of eumelanin are present in all brown animals; however in darker brown animals with far greater pheomelanin content, that amount of eumelanin has proportionately less effect on the measurable A_{650}/A_{500} ratio.

There was no significant difference in the amount of melanin present in white and fawn phenotypes (Figure 2.7). This could be explained by factors that affect the visual perception of the colour, or a factor that influences the test. For example, miscellaneous artefacts such as dirt, grass and other dried plant matter will result in an increase of absorbance (Table 2.2). However, because this technique has been demonstrated to provide a good estimate of melanin content (Ozeki *et al.* 1996) and steps were taken to ensure that samples were clean, it is probable that the lack of clear distinction between white, light fawn and fawn is real. There were large differences in the amount of eumelanin observed in animals that were described as black (Figure 2.7). This suggests that subjective assignment to colour is not always accurate, and/or that there may be multiple genetic and environmental pathways, which result in a black or diluted black colour.

The difference observed in the total melanin content between black and other colours in alpacas has also been well described in mice and sheep (Ozeki *et al.* 1995; Ozeki *et al.* 1996). Similarly the levels and types of melanins present in white alpacas are comparable to that of light coloured mice and sheep (Ozeki et al., 1996) and the

amount in brown animals closely resemble that of red mouse, human and sheep hair (Ozeki et al., 1996).

The HPLC analysis demonstrated that alpacas tend to be predominantly either eumelanic, or pheomelanic (Figure 2.10). Rosegrey is the only colour to display a large amount of both types of melanins in its fibre, which fits the model proposed by Munyard (2011), in which rosegrey is a pattern over a base colour, in this example the base colour is brown (equivalent to bay in horses). HPLC analysis also confirmed that brown alpacas do tend to be predominantly pheomelanic, and that there is only minimal DHI derived eumelanin present in any alpaca within our sample group (Figures 5) as is seen in rodents (Wilczek *et al.* 1996). DHI derived eumelanins tended to be produced proportionally to DHICA eumelanins in predominantly eumelanic animals, though with much lower amounts being present (Figure 2.11). The HPLC results demonstrated that there is no significant difference in melanin amount or type between the white and light fawn phenotypes, though there was a significant difference seen between these two and the fawn phenotype.

No significant difference in uniformity in the A_{650}/A_{500} ratios of brown and black alpacas was found (Figures 2.7, 2.8). This indicates that in alpacas of these colours, it is valid to compare the A_{650}/A_{500} ratios of samples from different sites. Similar "grid collection" techniques for fibre uniformity analysis have been used to determine the quality of fibre in different locations of the alpaca body (Aylan-Parker & McGregor 2002; McGregor *et al.* 2011).

However, in both of these colour groups the samples taken from behind the elbow had significantly lower A_{500} /mg levels than in other positions (Figures 2.10, 2.11). This indicates that comparisons of A_{500} /mg levels of samples collected behind the elbow cannot be compared directly to samples collected from other areas; however the amount of pigment present is still proportional to the intensity of the colour. For example in the brown group, the animals with greater average A_{500} /mg levels were still higher at the "behind the elbow" position than the animals with lesser average A_{500} /mg levels Therefore, the use of samples from behind the elbow is representative of overall pigment.

For the white phenotypes the A_{500}/mg levels showed no significant difference between different sampling positions (Figure 2.6). There was, however, a significant difference in the A_{650}/A_{500} ratios between the sample taken from the neck, which had a greater A_{650}/A_{500} ratio, than the samples taken from behind the elbow and shoulder (Figure 2.1). This may have been caused by environmental contamination, which may not be completely removed by washing. Another explanation may be that some animals were incorrectly classified as white, that were very pale patterned (e.g. tuxedo, grey or roan on a very light fawn base colour).

Overall, these results are the first to demonstrate that most brown, fawn and white alpacas are predominantly pheomelanic. This is an objective method for determining alpaca fibre colour, which has demonstrated that there is little difference in the type of melanin present in the fibre between many alpaca colour phenotypes. It is apparent that variations in pheomelanin levels are causing the changes in colour observed in these predominantly pheomelanic alpacas and therefore it seems most likely that a gene or genes affecting pheomelanin production (such as various *asip* alleles, or alternately a dilution factor) may be causing colour variations. Given that the vast majority of alpacas described as brown are likely to be predominantly pheomelanic, an amendment in alpaca colour terminology may be needed to more accurately reflect the genotypes present, to more closely align them with other species.

3

Colour Gene Expression in the Skin Transcriptome of White, Brown and Black Alpacas

"Technology gives us power, but it does not and cannot tell us how to use that power. Thanks to technology, we can instantly communicate across the world, but it still doesn't help us know what to say."

Jonathan Sacks

3.1 Introduction

There have been only a handful of studies looking at candidate genes for, and the mechanisms of, pigmentation in alpacas (Powell *et al.* 2008; Feeley & Munyard 2009; Cransberg & Munyard 2011; Feeley *et al.* 2011; Cransberg *et al.* 2013; Chandramohan *et al.* 2013). These data, combined with pedigree data, have enabled the development of a model of inheritance for alpaca pigmentation (Munyard 2011). However, there are still many colours and patterns in alpacas where the molecular basis of that colour/pattern is not understood, such as colour dilution, and patterns such as grey and roan. In this experiment, we began the search for as yet unidentified colour determining genes in the alpaca skin transcriptome. Further knowledge about which genes and variations affect colour will enable alpaca breeders to plan matings more effectively in relation to fibre colour.

In this pilot study RNAseq was used to analyse the transcriptome of the skin of white, bay and black alpacas. Melanocytes represent only a small amount of the total cells in mammalian skin – this ratio is approximately 36 keratinocytes to one melanocyte in humans (Fitzpatrick *et al.* 1967). This work has been used to validate whether RNA sequencing is an appropriate tool to study colour genes in the alpaca skin transcriptome. These colours were chosen to provide coverage of all genes expressed in eumelanogenesis and pheomelanogenesis. The white pool was chosen in order to compare the levels of expression of genes in animals with low and high levels of pigmentation.

3.2 Materials and Methods

3.2.1 Samples

Two-millimetre skin punch biopsies were collected from the left shoulder of 20 white, 20 bay and 5 black alpacas (Curtin University Animal Ethics Committee Approval Number N43-09). The animals were all housed on farms in the Perth region, and all were sampled within the same month. The black animals were originally collected as test samples for the extraction protocol, however after demonstrating the low levels of eumelanin production in bay skin (Cransberg *et al.* 2013), they were included in the RNAseq to ensure that full transcripts were obtained for genes involved in eumelanogenesis. Samples were stored in RNAlater (Ambion) at -20°C for up to two months prior to RNA extraction.

3.2.2 RNA extraction

Each tissue sample was homogenised in Trizol (Invitrogen) using the Fastprep system (Thermo Life Sciences) using lysing matrix D beads, according to the manufacturer's instructions. After room temperature incubation for 5 minutes, 2-bromochlorophenol was added and the samples were mixed. The aqueous phase was isolated by centrifugation. Isopropanol was added to the isolated aqueous phase to precipitate the RNA, which then was collected by centrifugation. The RNA pellet was washed once in 75% ethanol before resuspension in 100µl of RLT buffer (Qiagen RNAeasy kit). The RNAeasy kit was used to further purify the RNA according to the manufacturer's protocol.

3.2.3 RNA quality

The quality of the isolated RNA was assessed using the RNA pico 6000 kit on a Bioanalyser (Agilent). All samples analysed had an RIN >7.0. RNA was quantified using absorbance at 260nm on a Nanodrop spectrophotometer. Three separate pools of RNA were prepared by pooling equal amounts of RNA from all samples within each skin colour group.

3.2.4 Library Creation

Sequencing libraries were prepared from each of the three RNA pools using the Illumina TruSeq RNA sequencing library preparation kit, with a different TruSeq index used for each pool, according to the manufacturer's protocol. Each sequencing library was analysed for DNA size and purity using the Agilent Bioanalyser DNA 1000 electrophoresis chip. The quantity of DNA in each library was assessed using the Kapa Biosystems Library Quantification kit, as per the manufacturer's instructions.

3.2.5 Sequencing

The sequencing was performed on an Illumina Genome Analyser IIx with a standard paired-end sequencing flow cell. The flow cell was prepared using an Illumina Cluster station and a paired-end cluster generation kit. The three sequencing libraries were pooled in equimolar concentrations to a final concentration of 8pM DNA. The libraries were loaded into a single lane on the flow cell, with the other lanes being occupied by unrelated experiments. The flow cell was sequenced using the sequencing control program SCS2.6 and subsequently with RTA 1.6 for analysis. The reaction ran for 54 cycles, and was paired-end.

3.2.6 RNAseq analysis

An outline of the analysis of the RNAseq raw reads is shown in Figure 3.1.

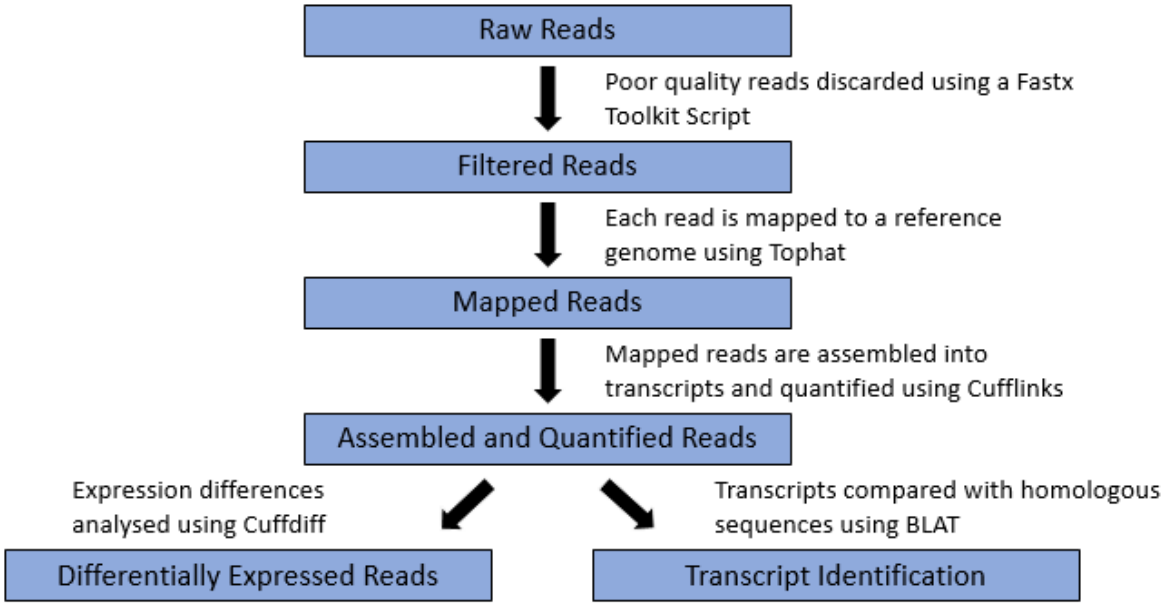

Figure 3.1. Simplified flow chart showing the RNAseq analysis pipeline.

3.2.7 Illumina sample read dataset preparation and quality filtering

Sequence data was de-multiplexed, separating reads according to their TruSeq index, and then converted into fastq format; (Illumina GenomeStudio 2009.2). Quality statistics were calculated and quality and nucleotide distribution plots were created using the fastq information script from the Fastx toolkit (Gordon & Hannon 2010). The Fastq quality filter script was used to discard reads where less than 90% of the sequence had a phred score of greater than 20. FastQC was used to check for read quality and it was determined that no trimming was necessary as there was no significant decline in read quality toward the end of the read.

3.2.8 Sample transcript alignment and assembly

Tophat 1.2.0 (Trapnell *et al.* 2009) was used to map the paired-end RNAseq reads to the vicPac1 (GCA_000164845.1) version of the alpaca genome and to the bovine bosTau8 UMD_3.1 genome (GCA_000003055.3) using default parameters. Cufflinks 1.2.0 was used to assemble these transcripts and to calculate transcript abundance using default parameters (Trapnell *et al.* 2010). These steps were completed separately for each of the three different colour groups. Differential

expression analysis was performed using Cuffdiff with default parameters (Trapnell *et al.* 2010).

An unannotated approximately 22× coverage alpaca genome subsequently became available (a draft version of the vicPac2 genome; hereafter referred to as vicPac2_draft genome). Tophat 2.0.6 and Cufflinks 2.0.2 were used to map and assemble, respectively, the RNAseq reads to the unpublished alpaca genome, using default parameters; with the exception of only allowing reads to map to one place in the genome (Trapnell *et al.* 2009; Trapnell *et al.* 2010). These three new assemblies (one per colour group) were combined using Cuffmerge to create a preliminary annotation of the genome. These predicted mRNAs were visualised using Geneious 5.5 (Biomatters) and the transcripts were exported. BLAT v34 (Kent 2002) was used to align the transcripts to a combined human (GCF_000001405.22), mouse (GCF_000001635.21) and cow (GCF_000003055.4; GenBank RNA database) RNA collection with the minimum accepted percentage homology by sequence set to 85%, and best match selected. This annotation, named vicPac2_draft_annotation_1, was used as the reference for the subsequent full-scale transcriptome analysis experiment (Chapter 4).

For the purposes of the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis, these transcripts were mapped via BLAT v34 to the human mRNA database (GCF_000001405.22; Kent 2002), with minimum accepted percentage homology by sequence set to 85%, and best match selected. These transcripts were then uploaded to The Database for Visualisation and Integrated Discovery (DAVID) for KEGG pathway analysis (Huang *et al.* 2009a, b).

The analysis described was repeated after the release of the vicPac 2.0 genome using the same conditions described above, however with updated versions of Tophat (v2.1.1), Cufflinks (v2.2.1) and BLAT (v36). This annotation of the vicPac2 genome was called vicPac2_annotation_1.

3.2.9 MART1 promoter analysis

3.2.9.1 Sample Collection

Blood was collected from 21 alpacas from farms across outer-metropolitan Perth, Western Australia. These included 8 black, 8 white and 5 bay individuals. DNA was extracted from the blood using the salt precipitation method described by Miller *et al.* (1988) or using a DNeasy blood and tissue DNA extraction kit (Qiagen), according to the manufacturer's instructions.

3.2.9.2 PCR amplification

PCR primers were designed using information from the Ensemble alpaca genome sequence (Genome assembly: vicPac1) to amplify the region 1.7Kb upstream of MART1 exon one (MART1L: AAAAACAGAGGCCACAATGAG; MART1R: CTTCAGCTGTGATGTGGGAG). Genomic DNA (50-100ng) was used as template for PCR using 2µM of each primer, 1× MyTaq reaction buffer and 0.75U MyTaq polymerase (Bioline) in a 10µL reaction. Amplification conditions were: 95°C for 2 min; 30 cycles of 95°C for 20s, 58°C for 30s and 72°C for 1 min 45s; then 72°C for 5 min.

3.2.9.3 Sequencing

PCR products were amplified from genomic DNA in five separate 10µL reactions and pooled prior to sequencing to eliminate bias due to PCR errors. Internal sequencing primers were used as well as the PCR primers to sequence the whole 1.7Kb amplicon (MART1iR: TTTCTCTTCTCTCCTCGCTG; MART1iL: AGACCTTTGGGATTTCATGAG). Sequencing was performed using the ABI Big Dye Terminator® system, and sequencing reactions were separated using a 48-capillary ABI 3730 DNA analyser (SABC, Murdoch). Geneious 5.5 (Biomatters) was used to visualise and analyse sequencing results.

3.3 Results

3.3.1 Sequence alignments

There were 25.4 million high quality paired-end 54bp reads generated. Of these 22.6 million mapped to the unannotated vicPac2 genome (Table 3.1). The percentage of reads that map to a genome are indicative, to an extent, of the quality of that genome. A comparison between the number of reads that map to the Vicpac_1.0 alpaca genome, a 2.5× coverage genome, and the number of reads that were able to map to vicPac2 (a 22x coverage genome), confirms that vicPac2 was an improved version (Table 3.1). The number of reads that map to genes in vicPac1 is sixfold less when compared to mapping the same reads to the genome of a relatively well annotated species such as the cow (BosTau8 genome), despite the genetic distance between alpacas and cows.

Table 3.1. Read statistics. The number of quality reads obtained and the amount of reads that mapped to different reference genomes and gene sets.

	White (Millions)	**Bay (Millions)**	**Black (Millions)**
Reads Obtained	9.4	8.9	7.4
Passed Quality Filter	9.3	8.8	7.3
Mapped to Vicpac1.0 Genome	2.1	2.0	1.6
Mapped to Vicpac1.0 Genome Genes	0.551	0.591	0.440
Mapped to Bostau8 Genome	5.1	4.4	2.2
Mapped to Bostau8 Genome Genes	3.9	3.6	2.7
Mapped to Vicpac2 Unannotated Scaffolds	8.5	7.7	6.4

3.3.2 Transcript analysis

The vicPac2_annotation_1 assembly included 29,308 unique transcripts (Table 3.2; full transcript list in Appendix 1). Of these, 126 transcripts were predicted to be keratin genes and keratin related genes (Data not shown).

Table 3.2. Assembly metrics of the transcripts created for each colour type and those when these files are merged.

	White	Bay	Black	Merged
Number of Transcripts	26,088	25,368	23,396	29,308
Minimum Transcript Length	53bp	54bp	54bp	53bp
Maximum Transcript Length	9057bp	9052bp	9058bp	9058bp
Mean Transcript Length	917bp	914bp	909bp	1040bp
Median Transcript Length	717bp	715bp	712bp	806bp
Total Length	23.9mb	23.2mb	21.3mb	30.5mb

Transcripts of many genes that are known to be involved in pigmentation were detected in this study (Table 3.3). Some of these genes are known to be expressed only in melanocytes in other mammals, for example *TYR* and *PMEL17*. In this dataset, *TYR* and *PMEL17* showed relatively high levels of expression, particularly in the black group, confirming that the skin biopsy sampling obtained melanocytes, and indicating that RNAseq is suitable for analysing expression of colour genes in the alpaca skin transcriptome.

Differential expression with at least a two-fold change was observed in eight of the 20 key colour genes focussed on between white, bay and black alpaca skin (Table 3.3). Some of these, like *MART1* and *PMEL17,* are expressed at low to undetectable levels in white compared to much higher levels in bay and black. Others are expressed at around the same level in all colours, (e.g. *MITF*, *KIT* and *KITLG*). Other genes, including the tyrosinase family, are highly expressed in black with lower expression in bay, and lower still in white; while *ASIP* is the only one of the key colour gene to show more than a two-fold increase in expression in white than in the other, more heavily pigmented, groups.

Table 3.3. Gene expression. Fragments per Kilobase of Transcript per Million Reads Mapped (FPKM) values calculated in alpaca skin for key colour genes.

		FPKM VALUES		
Gene	**Acronym**	**White**	**Bay**	**Black**
Agouti Signalling Protein	ASIP	95.8	14.8	1.6
Attractin	ATRN	7.9	7.4	6.7
B-Defensin 103	B-DEF 103	1.7	2.1	3.4
Dopachrome Tautomerase	DCT	4.7	7.0	42.7
V-Kit	KIT	4.5	5.7	6.2
Kit Ligand	KITL	49.0	42.2	46.8
Melanoma Antigen Recognized by T Cells 1	MART1	0.9	75.3	73.9
Membrane Associated Transport Protein	MATP	0.2	5.9	9.0
Melanocortin 1 Receptor	MC1R	0.8	7.4	10.1
Mahogutin	MGRN	5.5	5.6	4.7
Microphthalmia-Associated Transcription Factor	MITF	18.4	18.6	20.7
Melanoregulin	MREG	90.6	78.5	79.5
Myosin-5a	MYO5A	39.8	38.5	50.2
Palladin	PLDN	36.8	33.5	28.4
Rab27a	RAB27A	13.4	22.9	20.1
Premelanosomal Protein 17	PMEL17	0.0	195.0	186.4
Solute Carrier Family 36 A1	SLC36A1	7.2	9.3	9.2
Tyrosinase Related Protein 1	TYRP1	0.1	16.2	282.3
Tyrosinase	TYR	0.0	9.0	15.5

The coding sequence similarity of the selected alpaca colour genes, when compared to other well established and curated mammalian genomes such as cow, sheep and human, was very high. Generally, similarities at the nucleotide level were between 80-90%. The data in Table 3.4 demonstrates this for key pigmentation genes.

Table 3.4. Nucleotide sequence similarity (%) of key pigmentation genes compared between alpaca, cow and human. Results are based on a consensus of RNAseq reads over the CDS of the most common alpaca isoform of the gene.

Gene	**Cow**	**Human**
Agouti Signalling Protein	89	84
Attractin	90	87
B-Defensin 103	87	85
Dopachrome Tautomerase	87	84
V-Kit	89	86
Kit Ligand	89	88
Melanoma Antigen Recognized by T Cells 1	90	85
Membrane Associated Transport Protein	88	85
Melanocortin 1 Receptor	88	86
Mahogutin	89	85
Microphthalmia-Associated Transcription Factor	92	90
Melanoregulin	86	81
Myosin-5a	89	88
Palladin	86	85
Rab27a	82	81
Premelanosomal Protein	88	88
Solute Carrier Family 36 A1	91	88
Tyrosinase Related Protein 1	89	88
Tyrosinase	89	88

All of the key genes in the melanogenesis pathway were represented in at least two of the colour groups present (Table 3.3). Transcript analysis using the KEGG pathway demonstrates that all of the key genes involved in melanogenesis, with the exception of *α-MSH,* were detected in the skin (Figure 3.2).

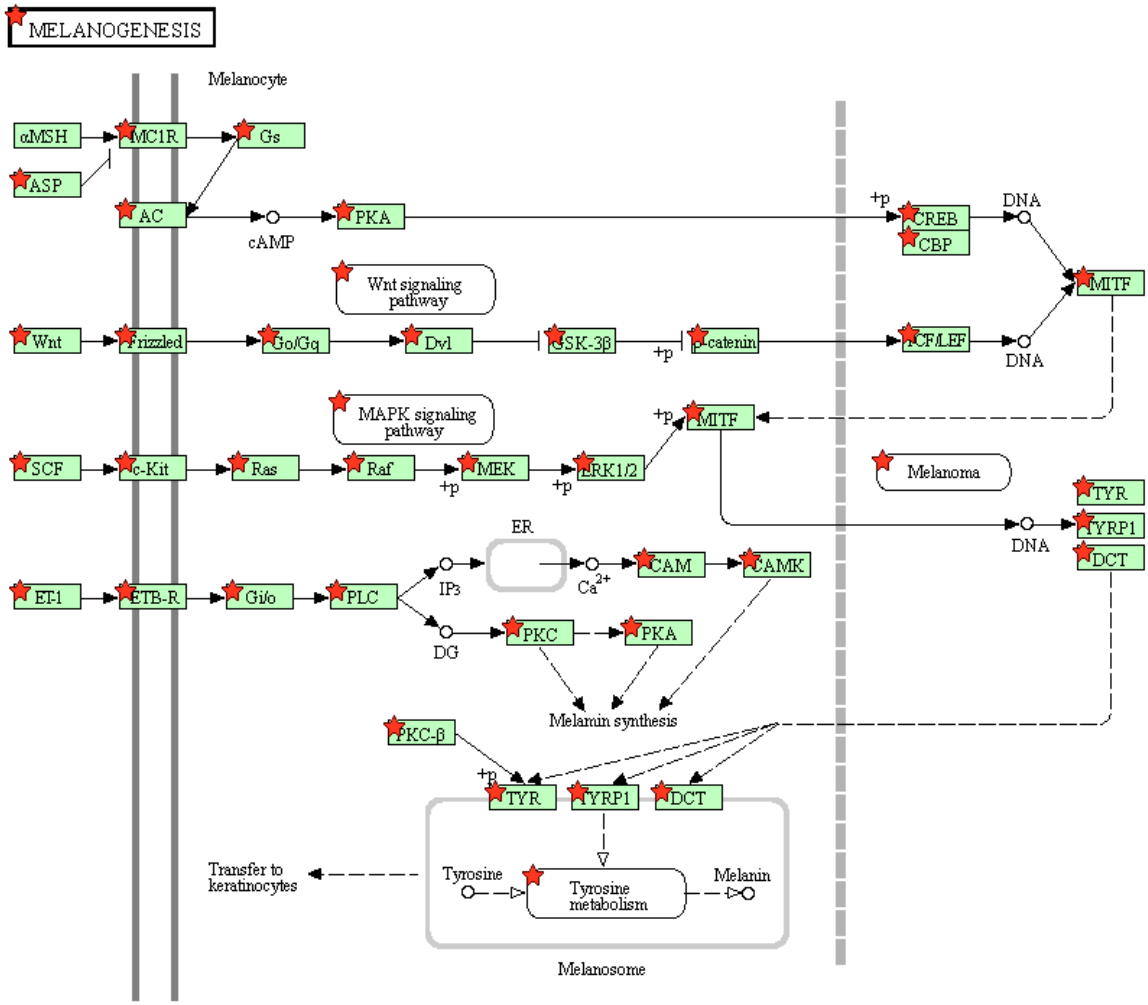

Figure 3.2. Melanogenesis KEGG pathway analysis based on the closest human homologs of the transcripts identified in the vicPac2_annotation_1 reference. Generated using DAVID (Huang *et al.* 2009a, b). All of the major colour genes were represented – with the exception of α-MSH, which is produced in the anterior pituitary.

Given the large difference in expression observed in *MART1* and *PMEL17* in white against bay and black animals, and that *MITF*, the key regulator of these genes was expressed at a relatively constant rate, the promoter region of *MART1* was analysed. MART1 has been demonstrated to be essential for PMEL17 expression (Hoashi *et al.* 2005). There were 11 SNPs identified in the region upstream from *MART1* exon 1 (Table 3.5). No statistically significant polymorphism associated with colour was found in the promoter region of *MART1* when the allele frequencies of the different colour groups were compared to a pool of all the SNPs found using a Pearson χ^2 test.

Table 3.5. SNPs found in the region immediately 5' of the *MART1* gene in alpaca. SNP nucleotide positions are representative of the number of bases 5' of the transcription start site. The Pearson χ^2 test was then used to investigate if an association existed between fibre colour and any of the SNP's found. The most significant P value for any colour type at each respective position is given below the SNPs. Statistical significance was accepted at $\alpha = 0.05$.

	539A/G	730C/T	812A/G	843A/G	929A/G	1006A/T	1086C/T	1191A/T	1191C/T	1327C/T	1482A/G
Black	G	C	A/G	G	A/G	A/T	C	A	C/T	C/T	G
Black	A/G	C	G	A/G	A/G	A/T	C	A/T	C/T	C/T	G
Black	A/G	C	G	A/G	A/G	A/T	C	A/T	C/T	C/T	G
Black	A/G	C	G	A/G	A/G	A/T	C	A/T	C/T	C/T	G
Black	A/G	C	G	A/G	A/G	A/T	C	A	C/T	C/T	G
Black	G	C	A	G	G	A	C	A	T	C	G
Black	A/G	C	G	A/G	A/G	A/T	C	A/T	C/T	C	G
Black	G	C/T	A/G	G	G	A	C	T	C	C/T	G
Bay	G	C	A	G	G	A	C	A	T	C	G
Bay	G	C	A	A	A	T	C	A	C	T	G
Bay	G	C	G	A	A	T	C	A	C	T	G
Bay	A/G	C	A/G	G	G	A	C	A	T	C	G
Bay	A/G	C	G	A/G	A/G	A/T	C	A	C/T	C/T	G
Bay	A/G	C	A/G	G	G	A	C	A	T	C	G
White	A/G	C	G	A/G	A/G	A/T	C/T	A	C/T	C/T	A/G
White	A/G	C	A/G	G	G	A	C	A	T	C	G
White	A/G	C	A	G	G	A	C	A	T	C	G
White	G	C	A	G	G	A	C	A	T	C	G
White	G	C	A	G	G	A	C	A	T	C	G
White	G	C	G	A	A	T	C	A	C	T	G
White	G	C	A/G	A/G	A/G	A/T	C/T	A/T	C/T	C	G
White	A/G	C	A/G	G	G	A	C	T	C	C	G
Min P-Value	0.82	0.56	0.35	0.24	0.13	0.13	0.67	0.15	0.14	0.07	0.83

3.4 Discussion

Many of the key genes involved in melanogenesis are highly conserved between species (Camacho-Hübner *et al.* 2002; Kumasaka *et al.* 2003). The high nucleotide similarity between the alpaca version of these key pigmentation genes and those of other species is more evidence of this conservation and supports the assumption that the functions of these genes are similar in alpacas to other species (Table 3.4). Furthermore, the expression levels of different colour-related genes observed in the skin of white, bay and black coloured alpacas agree with what is understood about pigmentation in other mammals (Slominski *et al.* 2004). For example, the lower expression levels of the tyrosinase gene family (including *TYR*, *TYRP1* and *DCT*) in white animals when compared to bay and further increase from bay to black (Table 3.3). The protein products of these genes are key to the production of melanin in mammals, with TYR being involved in the production of both eumelanin and pheomelanin, and TYRP1 and DCT being involved with the production of eumelanin (Kobayashi *et al.* 1995). Therefore it makes sense that such a marked decrease in *TYRP1* (17-fold; significant P = 0.005) and *DCT* (six-fold) expression was observed when comparing the predominantly pheomelanic bay colour group with predominantly eumelanic black colour group. Similarly there is also a large increase in *TYR* expression between different colour groups, from undetectable in white to expressed in bay (FPKM = 9.0) and black (FPKM = 15.5). This result is similar to the findings of Guibert *et al.* (2004) who demonstrated, in cattle, that the expression of *DCT* and *TYRP1* was completely suppressed in pheomelanogenesis, whereas *TYR* was only reduced. The increase of tyrosinase gene family expression from white to bay to black alpaca skin is complemented by the decrease of expression of the *ASIP* gene. It is well known that the interactions between *MC1R* and *ASIP* determine the type of pigment which is produced, so the nine-fold decrease in expression of *ASIP* between bay and black animals is expected, and may be the cause of the common expression patterns in the tyrosinase gene family. There is a further six-fold decrease in *ASIP* expression between bay and white. It has been demonstrated that the increased expression of agouti, through a gene duplication which puts it under the control of a constitutively expressed promoter, results in the white fibre colour in sheep (Norris & Whan 2008). The potential role of increased agouti expression resulting in diluted colour phenotypes is discussed further in Chapter 5.

Despite the small percentage of melanocytes in the skin of mammals, these data shows that RNAseq is an appropriate way of analysing gene expression in alpaca skin. All of the key genes in melanogenesis are expressed in at least two of the colours analysed (Table 3.3). This is supported by KEGG pathway analysis which shows that all genes involved in the melanogenesis pathway are expressed in the merged dataset (Figure 3.2). The exception to this is α-MSH, which is produced in the anterior pituitary and so is not expected to be present in the skin (Slominski *et al.* 2004). The level of expression detected in these colour genes, determined by FPKM values, meant that it was possible to observe detectable and biologically relevant differences in these key colour genes between colours.

The expression levels of genes encoding two enzymes involved in the production of mature melanosomes; *PMEL17* and *MART1* could also play a role in pigmentation (Hoashi *et al.* 2005). It has been demonstrated that PMEL17 fibres form a critical part of the amyloid like fibrils of the eumelanosome in human cells (Harper *et al.* 2008). The expression patterns of *PMEL17* and *MART1* in white, bay and black alpaca skin (Table 3.3) are comparable to what is seen in human melanoma cells, where decreases in *MART1* expression corresponded with decreased *PMEL17* expression (Hoashi *et al.* 2005). The low expression of *PMEL17* and *MART1* in white alpaca skin indicates that white alpacas may not be able to produce many, if any, mature eumelanosomes, thus severely limiting the amount of pigment able to be produced (Berson *et al.* 2001; Aydin *et al.* 2012). That there is a low number of mature melanocytes is supported by the large decrease in the expression of a number of melanosome specific genes, including *TYR*, in white animals. Without mature melanosomes the amount of pigment that is produced is minimal and thus the absence of mature melanosomes may be the cause of white fibre in alpaca, though how this absence is effected is still unknown.

The markedly lower expression of *PMEL17* and *MART1* in white animals as compared to bay and black animals prompted an examination of the *MART1* promoter sequence, to see if mutations in this region were causing the differential expression observed. In this region 11 SNPs were discovered (Table 3.5). None of the SNPs in the MART1 promoter region were found to correlate with colour.

Neither did any of the identified SNPs fall within key functional regions such as the distal or proximal promoters of *MART1*, the MITF binding site, nor in predicted miRNA binding sites (Butterfield *et al.* 1997; Du *et al.* 2003). One SNP, C1327T, was close to statistical significance in the black phenotype, but this is likely to be due to the small sample size, because both alleles that are observed in this position in black animals are also seen in the other two colours. Overall it seems likely that the SNPs in the region upstream of the alpaca *MART1* gene are not responsible for the changes in expression level that are observed in *MART1*, or for downstream *PMEL17* expression levels.

In conclusion, this experiment has provided insight into the genetics of pigmentation in alpacas. Given the conservation of the pigmentation pathway in mammals, it is reasonable to compare key colour genes from other species to those in alpaca, many of which appear to have similar functions when their expression is considered in lightly pigmented as well as pheomelanic and eumelanic animals. The very low expression of *MART1* and *PMEL17* in white animals could not be explained by mutations in the *MART1* promoter, but the low expression of these genes is likely to indicate the absence of mature melanocytes in white alpacas. The key regulating factors behind the changes of expression in these and a number of other key colour genes is yet to be elucidated, though it does not appear to be linked to the expression of *MITF*.

This study has demonstrated that RNAseq is a tool that can be used for the analysis of colour genes in the alpaca skin transcriptome. Furthermore, this study has enabled a preliminary annotation of the alpaca skin transcriptome, which will be used as a reference for future experiments.

4

The Alpaca Skin Transcriptome

It isn't the mountain ahead that wears you out; it's the grain of sand in your shoe.

Robert W. Service

4.1 Introduction

Coat colour in mammals is likely to be the result of interactions between a number of different genes, with polymorphisms found in *MC1R* and *ASIP* having already been linked to playing a role in coat colour phenotype in alpacas. However, these variations do not explain many subtle variations in fibre colour, nor various patterns observed in alpacas.

Analysis of the skin transcriptome of different coloured alpacas represents a chance to observe differences in the expression levels and coding sequence in all recognised and unrecognised colour genes, which can help us to further understand alpaca colour genetics. In the previous chapter it was demonstrated that RNAseq is sensitive enough to measure changes in the expression of melanocyte-specific colour genes from alpaca skin punch biopsies. These data also facilitated the development of a reference transcriptome and annotation. Subsequently, samples from a larger cohort of animals, representing eight colour phenotypes, were used to generate a more comprehensive alpaca skin transcriptome based on sequencing of ribo-depleted total RNA, on a SOLiD sequencing platform.

4.2 Methods

4.2.1 Sample collection

Three-millimetre skin punch biopsies were collected from the left shoulder of ten adult alpacas from each of the white, light-skinned fawn, dark-skinned fawn, light bay, bay, roan, silvergrey and black alpaca colour phenotypes (Curtin University Animal Ethics Committee Approval Number N43-09). Samples were stored in RNAlater (Ambion) on ice for three hours or less, before freezing at -80°C. The ten white samples were all from animals that were part of the cohort used in the pilot experiment (Chapter 3).

4.2.2 RNAseq

RNA was extracted using the RNeasy fibrous tissue mini kit (Qiagen) according to the manufacturer's instructions. A TissueLyser II (Qiagen) was utilised to aid homogenisation of samples. The quantity and quality of the extracted RNA was determined using an Agilent 2100 Bioanalyser and RNA 6000 pico kit (Agilent Technologies). Only samples with an RIN of 7 or higher were used for sequencing. Equal quantities of RNA from each animal of each colour phenotype were added to give an approximately 500ng total pool of RNA per colour type. The integrity of the RNA of each pool was then assessed using an Agilent 2100 Bioanalyser and RNA 6000 pico kit (Agilent Technologies), with each pool having an RIN between 7.5-7.7. The pools of RNA were then ribo-depleted using the Eukaryote RiboMinus Kit (Invitrogen) according to the manufacturer's instructions. Samples were then fragmented using RNase III and the fragments were purified using the RiboMinus Concentration Module (Invitrogen) as per the standard instructions.

Fragments were then assessed for appropriate size distribution, looking for a peak at approximately 200bp in size, using an Agilent 2100 Bioanalyser. Libraries were prepared for sequencing on an Applied Biosystems 5500xl (SOLiD) using the manufacturer's instructions (Applied Biosystems, Life Technologies). Briefly, SOLiD adapters were ligated to the RNA fragments. These were then reverse transcribed using the SOLiD Total RNAseq kit (Applied Biosystems, Life Technologies) and purified using the MinElute PCR Purification Kit (Qiagen). The

cDNAs obtained were then electrophoresed on a Novex 6% TBE-urea precast gel after which cDNAs between 150-250bp in length were excised from the gel. The excised cDNA was then amplified using the following conditions: 95°C for 5mins; 15 cycles of 95°C for 30s, 62°C for 30s and 72°C for 30s; and a final extension of 72°C for 7 min. Individual barcodes from the SOLiD RNA Barcoding Kit were used for each pool. Fragments were then purified using the PureLink PCR Micro Kit (Invitrogen). After confirming the library quality and size distribution using an Agilent 2100 Bioanalyser, the library from each colour group was applied to the SOLiD EZ Bead System, and then deposited across all six lanes of a flow cell and subjected to paired-end sequencing (75bp forward reads, 35bp reverse reads) by the SOLiD 5500xl system (Applied Biosystems, Life Technologies).

4.2.3 LifeScope software analysis

Whole transcriptome analysis was conducted using LifeScope v2.5.1 software with default conditions. The genomic reference sequence used for mapping was a draft of the vicPac2 genome (vicPac2_draft), which was the most complete reference available at the time. This was used in conjunction with the annotation file specifying the location of genic features and preliminary names of genes created during the previous transciptome experiment (vicPac2_draft_annotation_1, Chapter 3). LifeScope requires an annotated reference genome sequence for whole transcriptome analysis.

Following the initial analysis, a new reference annotation was created, to incorporate data from the much more extensive SOLiD experiment. Briefly, Samtools view v0.1.18 was used to convert the Binary Alignment Map (BAM) files representing two lanes worth of data from the LifeScope read mapping process to Sequence Alignment Map (SAM) files, printing the SAM header but otherwise using default options (Li *et al.* 2009). The output file was piped to the Picard tools SamToFastq utility v1.65 (https://sourceforge.net/projects/picard/) to convert it to fastq, using default options. These reads were then concatenated with the Illumina-generated mRNA fastq reads (Chapter 3) and analysed using the Tuxedo pathway (Trapnell *et al.* 2009; Trapnell *et al.* 2010). These reads from two sources could be mapped under the same conditions due to their similar fragment size. Tophat v2.0.14 was used to

align these reads to the vicPac2 genome and then the reads were assembled using Cufflinks v2.2.1. The Cufflinks gffread utility was used to output the transcripts in fasta format. To determine preliminary identifiers BLAT v36 (Kent 2002) was used to align these fasta transcripts to a combined human (GCF_000001405.33), mouse (GCF_000001635.24), cow (GCF_000003055.6), sheep (GCF_000298735.2), pig (GCF_000003025.5) and horse (GCF_000002305.2; GenBank RNA database), with the minimum accepted percentage homology by sequence set to 85%, and best match selected. This was supplemented with manual annotation of key colour genes (listed in Table 1.3) through creating a blast database of the vicPac2_draft genome using Geneious 5.5 (Biomatters) and then using blast to search against this database using mRNA from highly studied mammalian species with reference genomes (cow, human, mouse) or alpaca cDNA where possible, to find regions of similarity. In instances where there were multiple isoforms present, the most common isoform was used to blast against the scaffold database in order to find homologous sequences in the vicPac2_draft genome. The resultant gff annotation file, named vicPac2_draft_annotation_2 was used in a new reference mapping process in the LifeScope whole transcriptome analysis pathway which was repeated using default options. Heatmaps of gene expression were created using R v3.3.0 gplots libraries, heatmap2 function and by method h-clust.

4.2.4 Analysis using the Tuxedo suite of tools

NGS Plumbings xsq-convert utility v0.13.1 (http://pythonhosted.org/ngs_plumbing/index.html) was used to convert the raw xsq files into colorspace fasta (CSfasta) and qual (CSqual) format (using default options) and the resultant files were concatenated into full records for each colour group. Boxplots were drawn to visualise read qualities in Galaxy (Appendix 2; Blankenberg *et al.* 2010). Reads were then aligned to the vicPac2 genome using Tophat v2.1.1 with options changed to map using bowtie v1.1.2 (bowtie2 being incompatible with colorspace reads), changing the maximum number of hits per read to one, and specifying options to run using the CSfasta and CSqual files, using the combined SOLiD-Illumina annotation described above as a reference. These were assembled using Cufflinks v2.2.1 again using vicPac2_annotation_1 as a reference. The fastq reads extracted from the LifeScope mapped BAM files (described above) were then

also mapped and assembled to the vicPac2 genome, using Tophat v2.1.1 and Cufflinks v2.2.1 and default options, other than changing the maximum number of hits per read to one. Cuffmerge was then used to merge the gffs from all 8 colour groups (16 total) with the vicPac2_annotation_1 file, which resulted in an improved annotation named vicPac2_annotation_2 (Table 4.1 summarises these annotations).

Table 4.1 Summary of the different transcriptome annotations described

Annotation Name	Source
vicPac2_draft_annotation_1	Based on Illumina data
vicPac2_draft_annotation_2	Made from a combination of Illumina data and SOLiD data extracted from LifeScope BAM alignments
vicPac2_annotation_1	Based on Illumina data (vicPac2 genome)
vicPac2_annotation_2	Made from a combination of Illumina data, SOLiD data extracted from LifeScope BAM alignments and raw SOLiD reads (vicPac2 genome)

4.2.5 Comparison of Reference Genomes

The MUMmer3 software package (Kurtz *et al.* 2004) was used to assess the difference between the vicPac2.0 genome and the vicPac2_draft genome version which was utilised in the LifeScope analysis. First NUCmer v3.1 was utilised to align the two sets of scaffolds using default options. The alignments were then filtered to only include alignments between scaffolds that were over 10kb long and with over 95% homology. Mummerplot v3.5 was then used to generate a plot of these alignments with the sequences sorted by the 'best' alignment using the '-FAT' option of the unpublished draft version.

Tophat was used to confirm the effect of the two different genomes on the total number of reads mapping in an RNAseq experiment. Illumina RNAseq reads (Chapter 3) derived from pools of white, bay and black animals were mapped using Tophat v2.1.1 to the vicPac2 and vicPac2_draft genome, in both cases using standard options, but with reads only able to map in one place, no reference annotation was specified.

4.2.6 Transcriptome GWAS

A GWAS study was performed to analyse SNPs in the transcriptome. Firstly, SOLiD BAM files were sorted and indexed using Samtools v0.1.18 sort and index with default options (Li *et al.* 2009). Samtools mpileup was then used to call SNPs and indels from the sorted BAM files, setting the output to compressed bcf and computing genotype likelihoods (Li 2011). This output was then piped into bcftools view v0.1.18 to output a human readable vcf file. This file was then sorted and filtered by a script (C. Wade, unpublished, see Appendix 3) which filtered out SNPs detected with a coverage of less than 20, as well as calculating statistics on the raw coverage, coverage matching the reference and variant sequences, and the percentage of coverage matching to the alternative sequence. SNPs with the highest variation in allele frequencies between black and white animals were identified and there genomic location analysed for their proximity to known colour determining genes.

4.3 Results and Discussion

Approximately 50% of fragments mapped concordantly and in total 70% - 80% fragments mapped using the LifeScope software (Table 4.2). It would have been preferable to repeat the analysis with the release of the vicPac2 genome, as was possible with the Illumina data presented in chapter 3. However, by the time the official version of the vicPac2 reference was released, the server which had run the LifeScope software as described had been decommissioned, the LifeScope software was no longer supported, and was not able to be used. Therefore, the Tuxedo pipeline was used to reanalyse the data using the official vicPac2 genome. While these tools are compatible with SOLiD data, it became apparent that a great deal of optimisation must occur in the LifeScope software in order to achieve the high levels of mapping which were observed when utilising the LifeScope software with SOLiD reads. Concordant read mapping using default Tophat options was between 18-20% in all colours, with between 35-37% of total reads mapping.

Table 4.2 Alignment results from the eight colour groups analysed using the LifeScope software.

Colour	Raw Fragments (Millions)	Fragments Mapped	Fragments Mapped in Concordant Pairs	Total Fragments Mapped to Exons	Total Fragments Mapped to Introns	Total Fragments Mapped Intergenic	Unmapped Pairs and Low Map Records
White	117.9	71%	50%	16%	6%	29%	21%
Ls Fawn	137.5	75%	53%	18%	6%	30%	22%
Ds Fawn	184.2	75%	51%	15%	5%	32%	23%
Light Bay	280.5	75%	54%	18%	7%	31%	19%
Bay	265.4	79%	49%	16%	6%	29%	28%
Silvergrey	127.3	76%	53%	16%	6%	32%	22%
Roan	139.3	76%	52%	15%	5%	33%	24%
Black	161.1	83%	51%	16%	6%	35%	26%

In order to improve the read mapping several strategies were attempted, including analysing the data without a reference, increasing the numbers of mismatches allowed from 2 to 3, and varying the read gap length and mate pair inner distance. However, no appreciable increase in the percentage of reads mapping was achieved,

the largest increase was caused by increasing the number of mismatches allowed from 2 to 3 which resulted in an approximately 2% increase in the number of reads mapping, however with no increase in the total numbers of reads mapping in correct pairs. To eliminate the possibility that read quality was contributing to low mapping rates, reads were trimmed using the solidtrimmer.py script (Pedersen 2011), varying the minimum quality allowed, the minimum number of bases needed for the read pair to be kept, the maximum number of uncalled bases allowed for a read to be kept, and trimming the ends of reads whose quality fell below a cutoff threshold with a variable moving window. All options failed to lead to any improvement on the total number of reads which were mapped, though percentage read mapping did increase by approximately 1-2% with severe trimming stringency. Boxplots which were generated showed that the initial quality of reads was high and any more stringent trimming would likely result in the loss of high quality reads (Appendix 2).

NUCmer analysis of the vicPac 2 genome, compared to its draft, revealed that most of the component sequences are present in both genomes, albeit rearranged into a completely different set of scaffolds (Figure 4.1). Given that most of the scaffolds which were not represented in the draft, but were represented in the released genome, were relatively small (the majority – approximately 75% of the 4169 scaffolds - between 2kb and 20kb) and that the number of rearrangements was also relatively minor considering the size of the genome, it was thought that the likelihood of getting significant changes to the RNAseq output was minor. To confirm these assumptions, Tophat v2.1.1 was used to map reads from the previous Illumina experiment to both the vicPac2 draft genome and vicPac2 release version (Table 4.3). These data confirmed that the vicPac2 release version was more complete. However, the difference was not large enough to justify using the reads aligned with Tophat, specifically read counts, in place of the more completely mapped LifeScope reads.

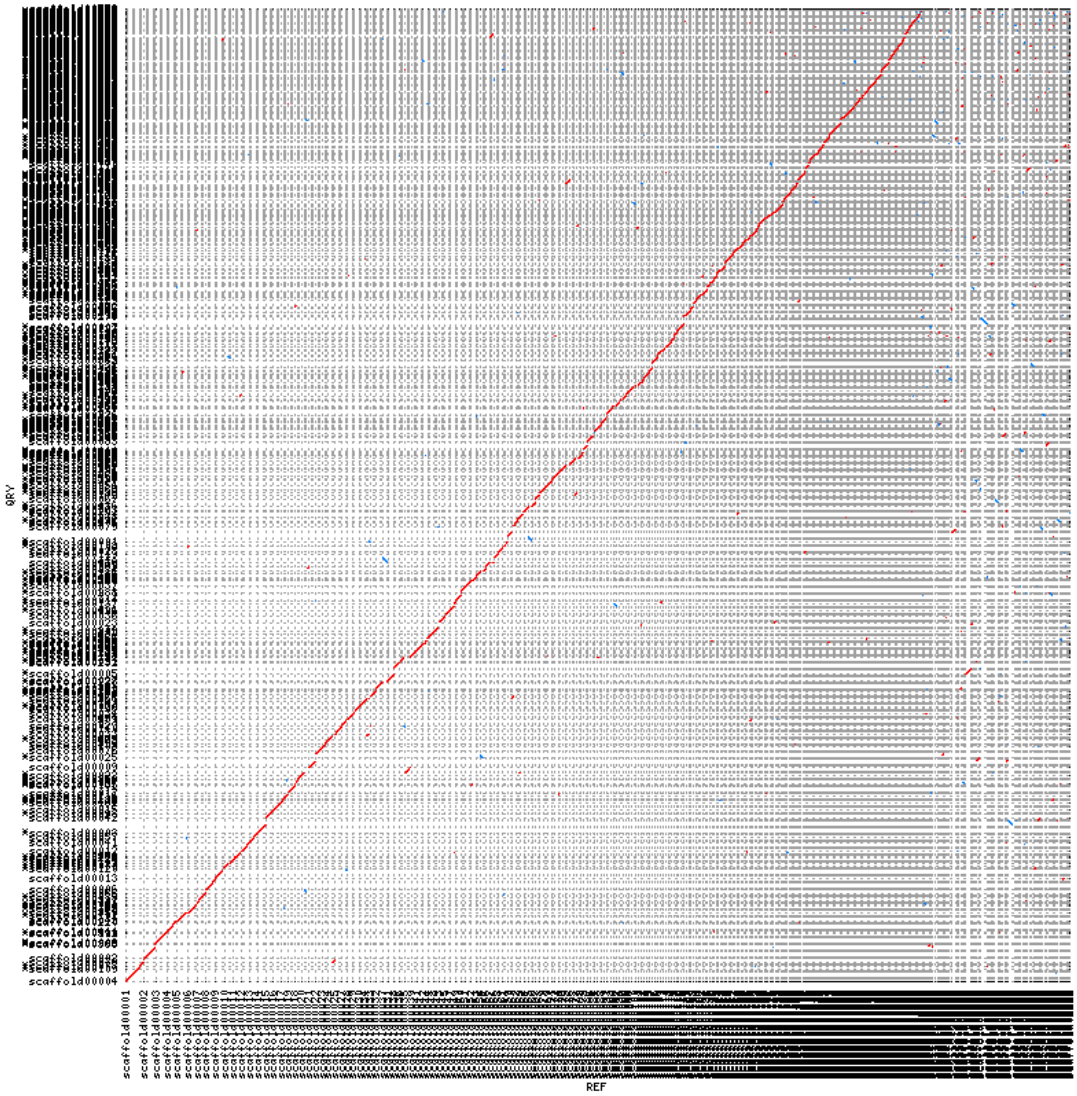

Figure 4.1. Sequence dotplot shows the alignment of the draft version of the vicPac2 genome (y-axis) and the final release (x-axis). Note that most of the sequence is similar between the two genomes, with some small rearrangements. The full size image is available in Appendix 4.

Table 4.3. Illumina RNAseq reads (Chapter 3) mapped to the vicPac2_draft genome vs. the vicPac2 release. Values are averages of the white, bay and black colour groups with error values representing one standard deviation.

	Unpublished Draft	**Vicpac2**
Total reads mapped	85.2%±0.8	91.2%±0.7
Multiple	0.9%±0.04	0.9%±0.04
Discordant	11.6%±0.2	12.1%±0.3
Concordant	72.8%±0.5	77.4%±0.7

The number of transcripts which were identified in this experiment was much larger than in the pilot study (Table 4.4; Table 3.2; full transcript list in Appendix 5). The major reasons for the increase is likely to be the rRNA depleted sequencing used, compared to the mRNA sequencing in the Illumina dataset; the increase in depth of coverage, with over 60-fold more reads sequenced; and potentially the increase in different colours analysed.

Table 4.4. Transcript metrics of vicPac2_annotation_2, including Illumina reads, LifeScope mapped SOLiD reads which were subsequently remapped using the Tuxedo pipeline, and raw SOLiD reads which were mapped in the Tuxedo pipeline without being mapped in LifeScope first. Coverage represents the coverage of transcripts over the vicPac2 genome.

Total Number of Transcripts	361,295
Average Transcript Length	761.8bp
Median Transcript Length	373bp
Maximum Transcript Length	36,445bp
Total Length of Transcripts	275.2Mb
Coverage	8.12%

Most of these reads map to intergenic regions (Table 4.2), which supports the case that many of these new transcripts will be genes that were not seen with the lower coverage of the Illumina sequencing and total-RNA constituents. One of the key features of these data was the presence of a large number of nascent (incompletely processed) RNA transcripts, which extended beyond the polyadenylation site. This has been described before in total-RNAseq data (Ameur *et al.* 2011). The presence of nascent RNAs did not affect the measurement of expression of the colour genes listed, as these were manually annotated. However, it should be noted that other transcripts may have nascent RNA present, which could affect the Reads Per Kilobase of Transcript Per Million Reads Mapped (RPKM) of the mature mRNA component of the gene, account for more isoforms caused by immature RNAs, and increase the fragment size – through having unprocessed 3' sequence. The example in Figure 4.2 shows the potential effect of nascent RNAs on the AHCY gene, with intron read-through potentially combining two exons (though this could also be incomplete splicing), and also potentially extending the 3' UTR past the length

detected by the Illumina study (which should be correct dependant on complete coverage of the gene). Nascent RNAs will also be the cause of a large portion of the reads mapping to introns and to intergenic regions, which will in-turn also increase the number of transcripts observed.

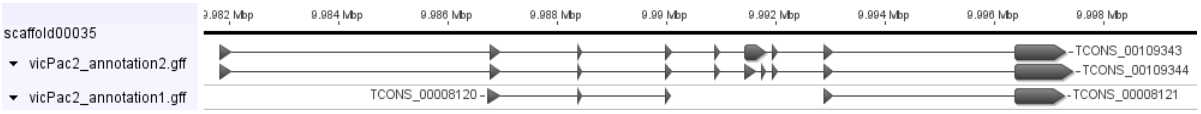

Figure 4.2 Example of improvements in the vicPac2_annotation2 compared to vicPac2 _annotation 1. Full gene structure of the AHCY gene observed in the vicPac2_annotation_2 vs. the partial gene structure, which was available in vicPac2_annotation_1.

The total number of transcripts found was much higher than that found in other RNAseq studies focussed on the skin of sheep (74,533 in black skin 90,006 in white skin; Fan *et al.* 2013) and cow (39,577; Weikard *et al.* 2013). A factor other than nascent RNA which will contribute to the increase in total transcripts is the presence of non-coding RNA. This is supported by the large proportion of smaller transcripts detected (Figure 4.3), with fragment lengths being smaller than what was reported in cow (Weikard *et al.* 2013). These smaller fragments are likely to include many different non-coding RNA species. There are also many unrecognised transcripts, which are likely to be a combination of previously undocumented RNAs and artefacts, with many transcripts sharing homology to unclassified RNAs, or from not finding matches against any homologous sequences in the database. The increased proportion of small RNAs also resulted in a decrease of average and median transcript lengths (Table 4.4) despite more complete coverage causing obvious increases in the completeness of the annotations of many other genes (Figure 4.2). Other factors which played a role in increasing the total number of transcripts include the fact that a greater variety of colours of animals were sequenced, which, coupled with the increased depth of sequencing, is likely to be the cause of more alternate isoforms being observed.

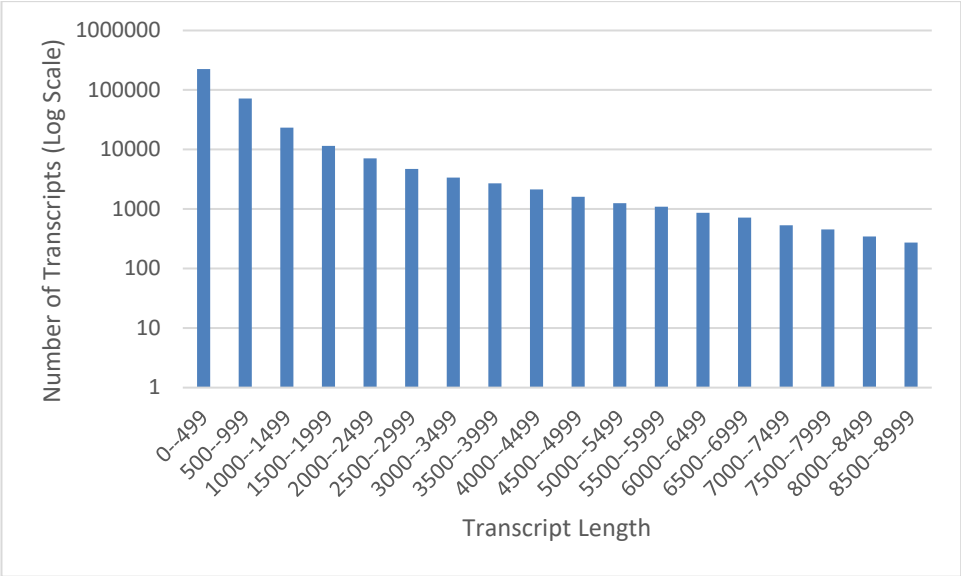

Figure 4.3. Log distribution of transcript size for the skin transcriptome annotation.

Similarly to what has been described in sheep (Fan *et al.* 2013) keratin genes make up a large proportion of the most highly expressed genes in alpaca skin, with 23 of the top 100 highly expressed genes in alpaca skin being keratin related. A heatmap showing the expression of the 50 highest expressed genes highlights that these different keratin genes have no correlation, in terms of their expression levels, with the different coloured alpacas (Figure 4.4). The expression of the 50 most highly expressed genes links light bay, bay and the light-skinned fawn colour groups. There are also interesting correlations between the colour groups in the expression of different Small nucleolar RNA, C/D box genes (SNORD, a snoRNA), which are some of the highest expressed genes that were found in this dataset. SNORD RNA species have been shown to act like miRNAs (Ender *et al.* 2008) as well as being the cause of alternate splicing events (Kishore & Stamm 2006). Previously miR-25 has been demonstrated to have a role in alpaca pigmentation through the regulation of *MITF* (Zhu *et al.* 2010). miR-25 was not found in this dataset, with the only area showing homology to the human and mouse miR-25 sequences not covered by any transcript. The differences in expression of these snoRNAs could be an artefact due to the tissue composition of the skin biopsy taken, but their functions are not completely understood (Ender *et al.* 2008) and it is possible that they play a role in alpaca pigmentation.

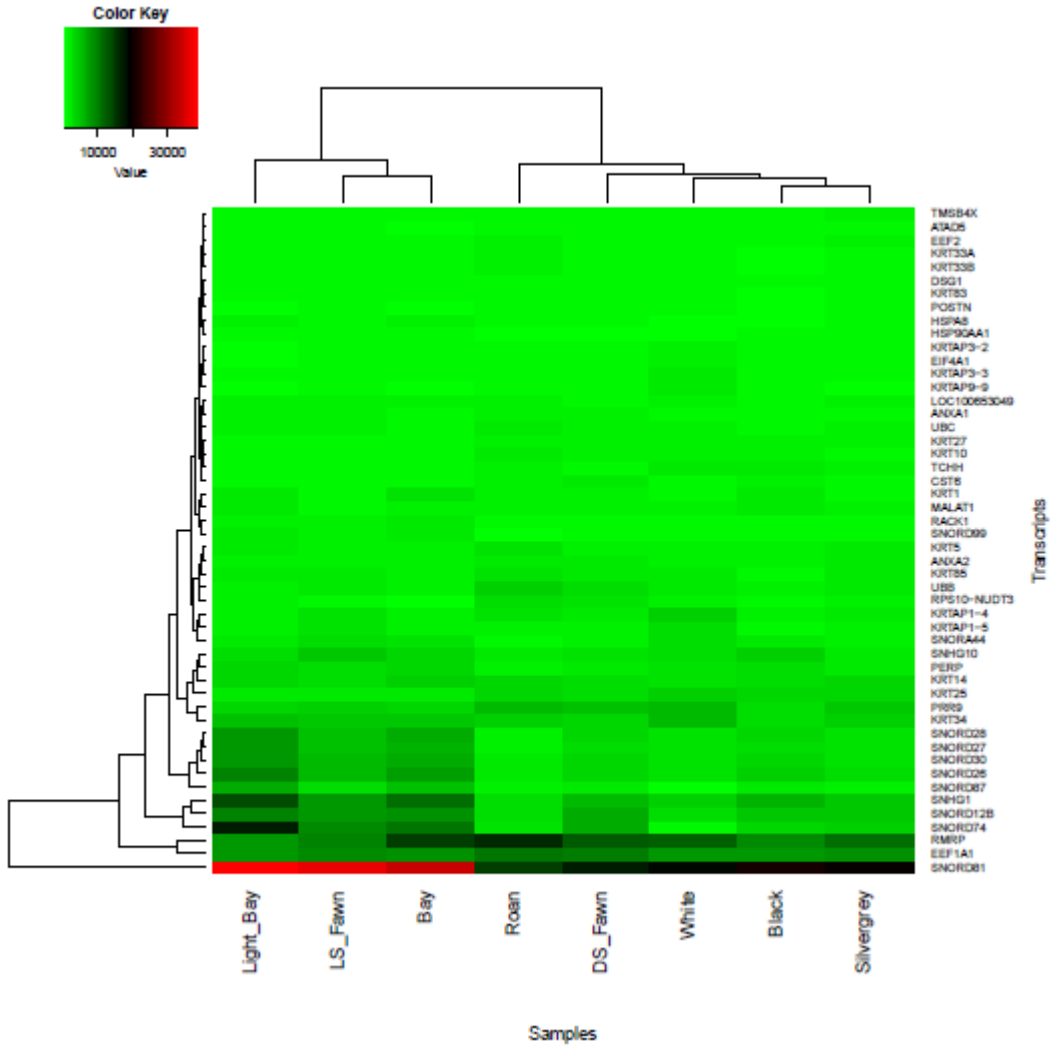

Figure 4.4. Heat map and hierarchical clustering of the 50 most highly expressed genes by FPKM, and the colour groups analysed.

The cellular components and biological functions of highly expressed genes varied in the alpaca skin transcriptome when interrogated using GO-terms. Genes associated with key skin processes, such as those involved with cell to cell adhesion and epidermis development, were well represented in the set of most highly expressed genes (Figure 4.5). Genes associated with key structural parts of the skin were also well represented, with components such as keratin filaments and desmosomes amongst the most enriched genes (Appendix 6). Cytoplasm and cytosol were the best represented of the cellular components (Figure 4.6). Interestingly, of the top 100 highest expressed genes most were associated with extracellular exosomes, a contrast to what is measured in sheep skin, where relatively few genes associated with extracellular processes were represented in a GO analysis of the whole transcriptome

(Fan *et al.* 2013). These extracellular exosomes are of interest from a pigmentation perspective as exosomes released from keratinocytes have been demonstrated to modulate pigmentation in humans by altering gene expression and enzyme activity (Cicero *et al.* 2015). They transport specific miRNAs such as miR-203 which was highly expressed in exosomes from black skin, and linked to an increase in TYR protein levels (Cicero *et al.* 2015). No homologous sequence to miR-203 was found in the alpaca genome.

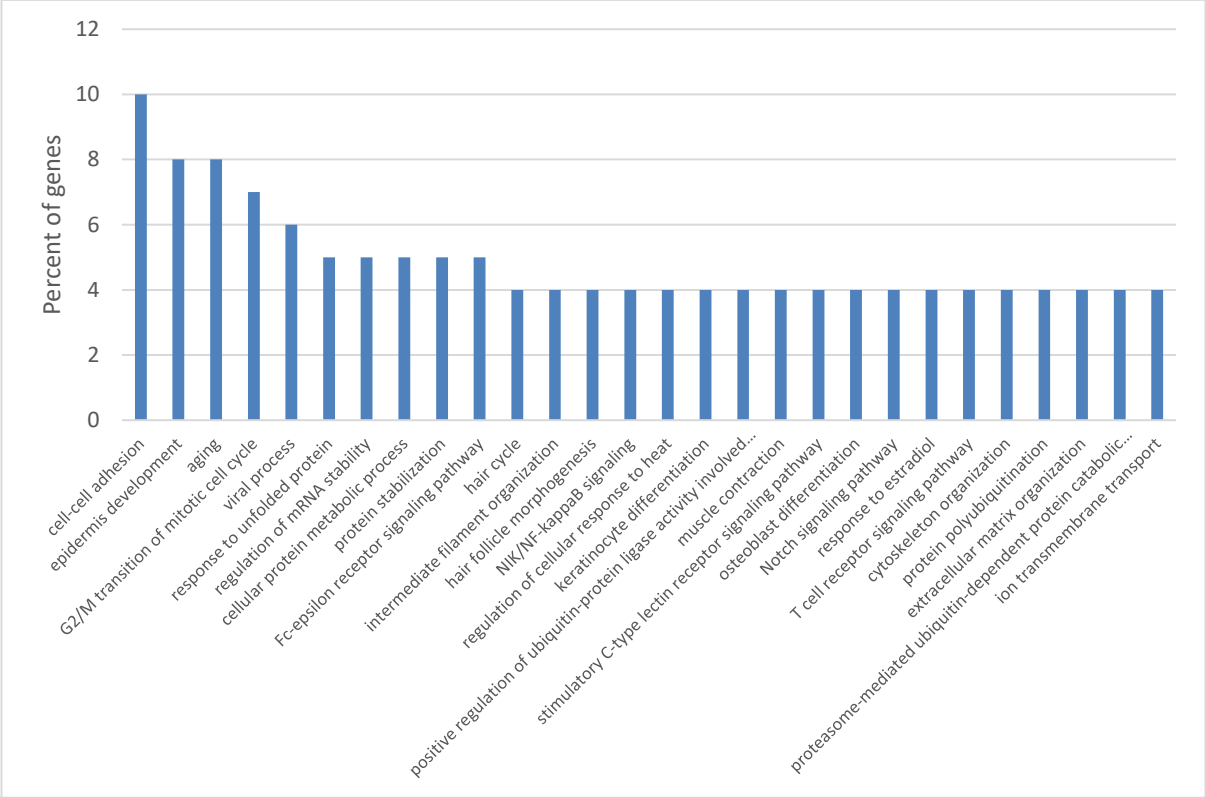

Figure 4.5. GO terms in the category biological process for the top 100 non rRNA expressed genes, Processes with less than 4% of genes associated were not included.

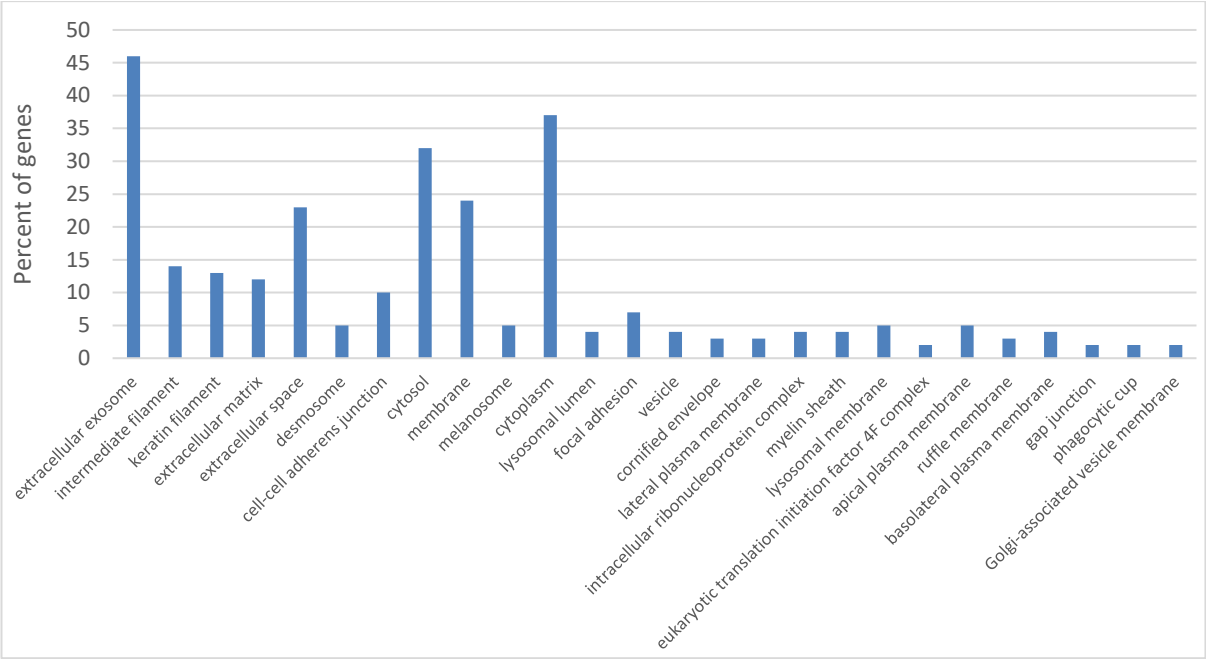

Figure 4.6. GO terms in the category "cellular components" for the most highly expressed 100 non rRNA genes

When analysing the expression of key colour genes, the predominantly pheomelanic colour groups are clustered together (Figure 4.7). The *TYRP1* gene appears to be the most distantly related gene as it clusters on a separate branch, but is likely to be due to the large fold changes in expression observed between different colour groups, in particular between the lighter colours (i.e. white and fawn) and darker colours (i.e. black; Table 4.5). The other most distant gene is *ASIP*, which is expected as its expression is markedly higher in the lighter colours, unique amongst the major colour genes analysed (Table 4.5). Genes known to be co-expressed such as *TYR* and *MATP* clustered closely together (Baxter & Pavan 2002) while other co-expressed colour genes were further apart than anticipated (e.g. *PMEL17* and *MART1*; Hoashi *et al.* 2005). *PMEL17* and *MART1*, along with *DCT*, showed a very similar gene expression profile across the different colour groups (Figure 4.7; the protein-protein interactions between these key colour genes are represented in Figure 4.8).

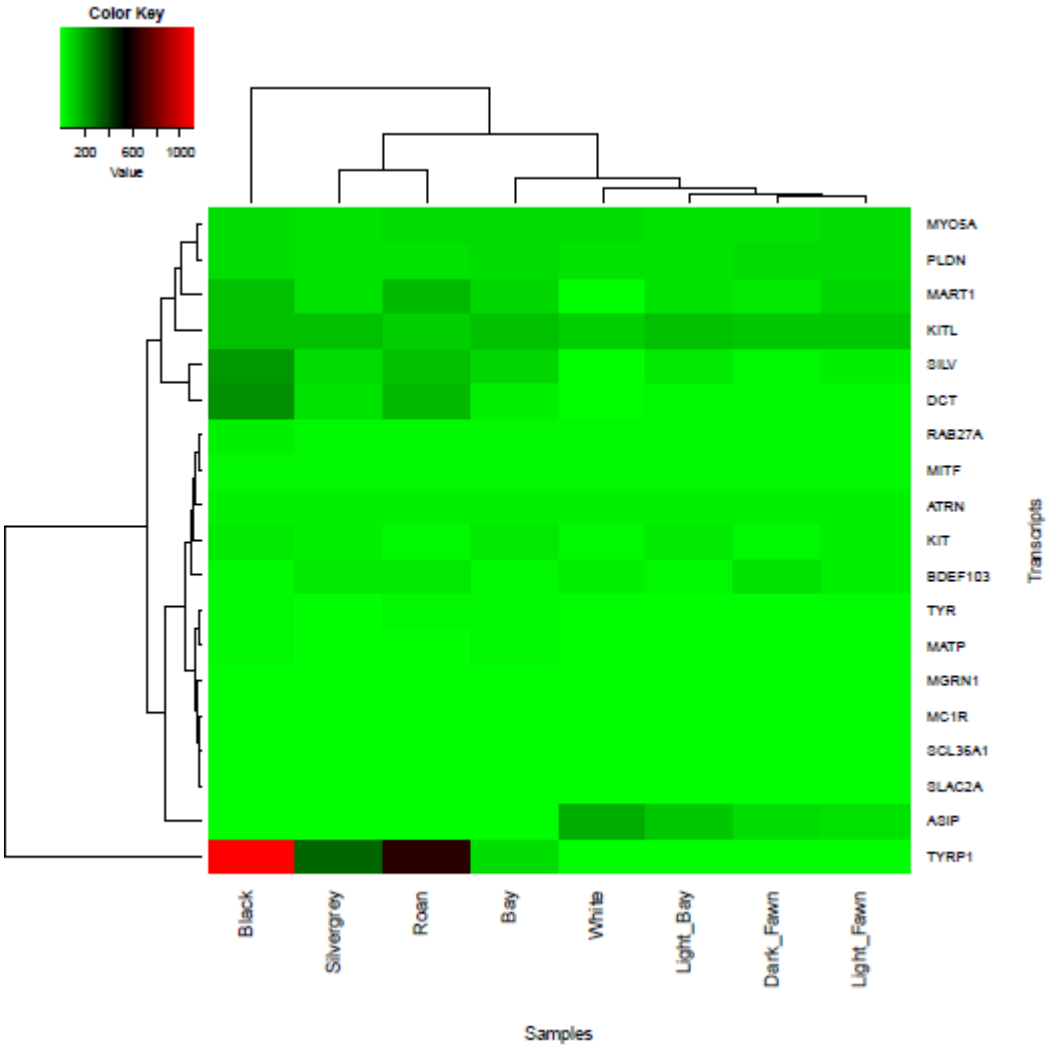

Figure 4.7. Heat map and hierarchal clustering of key colour gene expression and the colour groups analysed.

Table 4.5. Expression levels (FPKM values) in a selection of key colour genes, based on their role in determining fibre colour in mammals, across the 8 colour groups analysed.

	White	**LS Fawn**	**DS Fawn**	**Light Bay**	**Bay**	**Roan**	**Silvergrey**	**Black**
ASIP	187.85	71.37	80.1	123.41	9.49	1.62	3.81	0.46
ATRN	34.22	31.35	30.61	32.85	35.84	34.45	37.28	36.4
B-DEF 103	34.11	32.36	64.38	16.68	22.03	44.51	55.19	29.53
DCT	10.4	17.81	16.8	15.66	39.88	161.37	70.97	240.6
KIT	16.73	40.31	29.43	48.78	56.23	26.04	33.68	46.36
KITL	105.74	124.96	128.18	136.04	141.85	104.13	139.87	138.49
MART1	13.29	97.31	45.93	63.3	100.71	153.05	67.84	142.34
MATP	1.34	7.71	3.34	12.9	17.89	14.62	8.16	24.79
MC1R	0.33	5.46	1.72	3.27	2.29	5.6	4.78	7.27
MGRN1	9.41	9.4	7.35	6	6.51	7.49	12.43	12.74
MITF	18	21.82	20.68	24.08	28.81	29.47	22.06	27.95
MYO5A	74.44	77.57	72.66	70.35	82.1	78.74	73.58	74.82
PLDN	66.76	78.27	77.82	62.69	74.46	65.12	61.93	79.53
PMEL17	5.74	39.97	16.22	56.96	102.77	140.46	75.17	227.7
RAB27A	22.4	28.74	19.8	28.65	27.39	26.01	26.58	29.73
SCL36A1	1.5	2.53	0.75	1.17	2.4	3.54	2.01	3.09
SLAC2A	0.5	0.67	0.54	1.25	0.25	0.83	1.36	2.33
TYR	1.06	10.41	3.12	6.82	21.14	22.46	8.77	28.01
TYRP1	1.78	4.05	3.33	4.95	86.05	644.03	334.84	1103.77

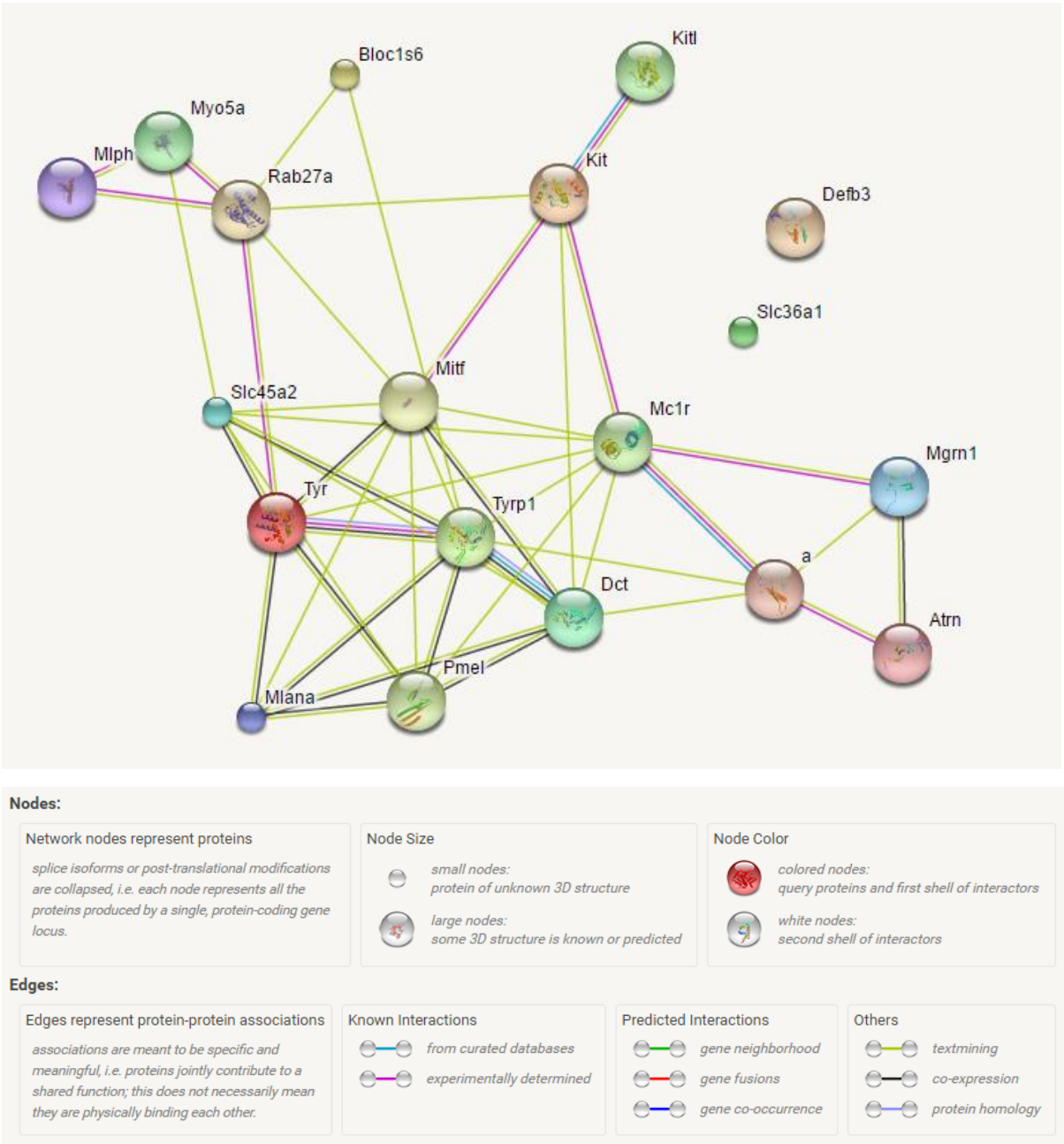

Figure 4.8 StringDB mouse protein interactions between the key colour genes analysed (Szklarczyk *et al.* 2014). Key: a=ASIP; bloc1s6=PLDN; mlph=SLAC2A; MLANA=MART1; defb3=βDEF103. Adapted from http://string-db.org/.

The light-skinned fawn pool of alpacas tended to have higher expression of several key colour genes than the dark-skinned fawn group. The major differences noted were a three-fold increase in *TYR* and *MC1R* as well as two fold increases in *PMEL17*, *MART1* and *MATP* expression (Table 4.5). Low levels of TYR are associated with pheomelanin production (Prota 1993; Ozeki *et al.* 1997a), however an increase in *TYR* can explain the increased pheomelanin production because *TYR* is required for pheomelanin synthesis (Kobayashi *et al.* 1995). These differences in

gene expression can be attributed to the selection of animals, where the light-skinned fawn group had several animals that had enough pheomelanin to be described as chestnut in their fibre, whereas the majority of the animals in the dark-skinned fawn group contained little pigment and were close to white in colour, so the upregulation of genes needed in melanocytes, particularly *MC1R* and *TYR* is logical.

In order to identify SNPs that are associated with colour a genome wide association study (GWAS) was performed on the black and white colour groups. GWAS can utilise RNAseq data, however the process is complicated because coverage will change with the expression of genes, which means that SNPs will not always be expressed at high enough coverage to confidently detect variants. However approximately 12% of SNPs found in GWAS studies are known to be found coding regions (Freedman *et al.* 2011; Costa *et al.* 2013). Two of the SNPs with the highest variation between the white and black pools, which were represented above the minimum coverage (20×), both mapped downstream of agouti (Table 4.6). The nearest of these SNPs is positioned approximately 14Kb downstream of *ASIP* exon 4, in the *AHCY* gene. The SNP lies at position 510 in the *AHCY* cds, but lies in the 3rd codon position and does not result in a nucleotide change. The other SNP lies approximately 165kb downstream of the *ASIP* gene 137bp into the 3' UTR of the *ITCH* gene. No other SNPs with large variation between the frequencies of alleles observed between the black and white colour groups mapped to regions containing known colour genes. These data further supports the hypothesis that the *ASIP* gene plays a role in determining colour, particularly between black and white animals, and the fact that there were no SNPs in the ASIP cds might suggest that the effect is one of regulation, rather than protein function. The variable expression of *ASIP* in different colour groups strongly suggests that it plays a major role in determining alpaca fibre colour, and Feeley *et al.* (2011) have already demonstrated the role that agouti alleles have in determining the black colour in alpacas. Potential further roles for agouti in determining alpaca fibre colour are discussed in Chapter 5.

Table 4.6. Top 10 SNPs which vary in frequency between the black and white colours. A minimum coverage of 20 was required. Genes used in the relative position column are the best hit found for the transcript via BLAT comparison with known RNA sequences, and exon position is based on the exon number of the transcript when aligned to the vicPac2 genome.

Relative Position	Synonymous (S) / Non Synonymous (Ns)	Allele Frequency in White (coverage)	Allele Frequency in Black (coverage)
Ring Finger Protein 180, Exon 4	S	0.00 (30)	0.79 (43)
S-Phase Kinase Associated Protein 1, Exon 5	S	0.01 (106)	0.72 (92)
Itchy E3 Ubiquitin Protein Ligase, Exon 12	Non-coding	0.78 (27)	0.08 (25)
BMP and Activin Membrane Bound Inhibitor, Exon 3	Non Coding	0.21 (28)	0.85 (39)
Metastasis Associated Lung Adenocarcinoma Long Non-Coding RNA, Exon 1	Non Coding	0.62 (24)	0.00 (134)
Alcohol Dehydrogenase 1 Like, Exon 3	NS Ile/Leu	0.19 (32)	0.81 (64)
Adenosylhomocysteinase, Exon 4	S	0.61 (28)	0.00 (21)
Unidentified RNA	N/A	0.00 (1864)	0.60 (1201)
Serpin Peptidase Inhibitor Clade B Member 5, Exon 2	NS	0.76 (72)	0.19 (47)
Nuclear Factor of Kappa Light Polypeptide Gene Enhancer in B-Cells Inhibitor Alpha, Exon 6	Non-Coding	0.77 (47)	0.20 (49)

In summary, the alpaca skin transcriptome was analysed using a combination of both LifeScope and open-source software. This resulted in a total of 361,295 unique transcripts, which included an unknown proportion of nascent RNA, and different splicing isoforms. Analysis of the most highly expressed genes demonstrated that, as expected, genes encoding keratins and desmosomes were enriched in skin. Hierarchal clustering of the 100 most highly expressed genes suggested a potential role in pigmentation for snoRNAs. A GWAS study between white and black colour groups highlighted the area of the genome containing *ASIP*, and evidence suggests that *ASIP* expression has a major role in determining alpaca fibre colour.

5

The ASIP Gene is Upregulated in White and Dilute Phenotypes in Alpacas

"A man should look for what is, and not for what he thinks should be."

Albert Einstein

5.1 Introduction

Cransberg *et al.* (2013; expanded in Chapter 2) showed that the ratio of eumelanin to pheomelanin in the base colour alpaca fibres was similar in animals that had white, light fawn to dark fawn; and light through to dark bay coat colours. This indicates that colour dilution genes are having an effect on alpaca fibre colour. The subtle variations that were seen, most notably in the levels of pheomelanin between different coloured animals, suggests that these colour differences may be caused by quantitative changes. Variation in gene expression can explain these quantitative changes, rather than altered protein function single point mutations and indels, which tend to have larger and/or qualitatively different effects on gene function, such as those previously described in alpacas (Powell *et al.* 2008; Feeley & Munyard 2009; Feeley *et al.* 2011).

In addition to its role in switching between eumelanin and pheomelanin production (Oyehaug *et al.* 2002), excess expression of *ASIP* has resulted in dilution, or complete absence of colour in mammals: In human melanocyte cultures the addition of ASIP resulted in decreases in both eumelanin and pheomelanin production (Le Pape *et al.* 2008). In sheep, an *ASIP* gene duplication puts it under the control of a constitutively expressed promoter, resulting in increased expression and a white phenotype (Norris & Whan 2008). Further, increased *ASIP* expression has been demonstrated to cause lighter phenotypes in mice, notably different dorsal/ventral expression resulting in the lighter underside of beach mice (Steiner *et al.* 2007; Manceau *et al.* 2011). These reports, and the data from these experiments support the hypothesis that increased *ASIP* expression is the cause of dilute colour phenotypes in alpacas. The *ASIP* gene has three coding exons, named exons 2, 3 and 4, which are preceded by a non-coding exon (or exons) in alpaca. In various species multiple isoforms of *ASIP* exist which vary only in their 5'UTR exons, and which in many cases result in altered gene expression (Bultman *et al.* 1992; Vrieling *et al.* 1994; Girardot *et al.* 2005; Drögemüller *et al.* 2006; Fontanesi *et al.* 2010; Chandramohan *et al.* 2013; Ciampolini *et al.* 2013) *ASIP* 5'UTR gene structures available in Appendix 7).

5.2 Methods

5.2.1 RNAseq

Please refer to Chapter 4.

5.2.2 Analysis of the *ASIP* 5' UTR

The roan and silvergrey samples are not analysed here as they were sequenced for an unrelated experiment, these data are included in chapter 4 to provide additional information about eumelanic animals, because both phenotypes have a black base colour.

BAM files that resulted from LifeScope mapping of SOLiD sequences were converted to SAM files using Samtools view v0.1.18 and subsequently from SAM to fastq using Picard tools SamToFastq v1.65 (Li *et al.* 2009; http://picard tools (http://picard.sourceforge.net), both using default settings.

A pool of sequences representing one sixth of the total reads for each colour (i.e. One flow cell worth of data as described in Chapter 4) were aligned to the exon 2 region of the alpaca *ASIP* gene (sequence extracted from the vicPac2 genome) using the Geneious aligner program v8.1.9 (Biomatters), with high stringency mapping (reads with more than 5% mismatches discarded; a step mapping process described in Figure 5.1; Step 1). Stringently aligned read pairs in which one read crossed the *ASIP* exon 2 5' boundary were used to extend the known 5' UTR region by adding this additional 5' unique sequence to the 5' end of the *ASIP* exon 2 sequence (Step 2). Where the aligned reads varied at the 5' end new consensus sequences were created, and reads were subsequently stringently remapped to these consensus sequences (Step 3). This process was repeated with each isoform thus "walking" the sequence out until no more reads could be found to overlap the 5' border of the contiguous sequence.

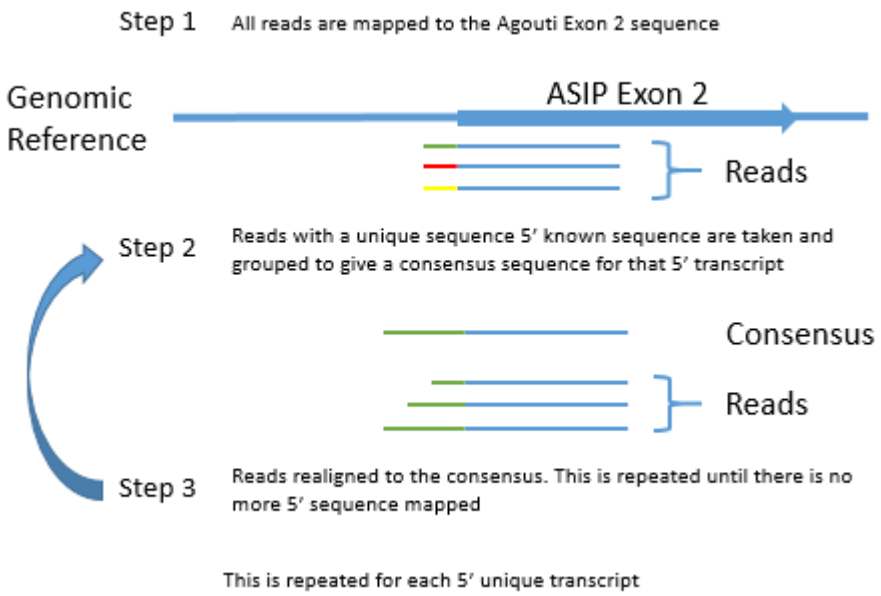

Figure 5.1. The iterative process shows the steps by which the sequence from each unique 5' UTR *ASIP* transcript was identified. Different 5' colours represent variant 5' sequences.

To identify the genomic positions of these 5' exons the consensus sequences were then aligned to a blast nucleotide database of the 4120 scaffolds comprising the vicPac2 genome using Geneious (Biomatters; prior to the release of the 2015 vicPac2 genome on NCBI).

An approximately 100KB region immediately upstream of *ASIP* exon 2 was amplified in a series of long range PCRs and was sequenced on the Ion Torrent (K. Munyard, Unpublished) to explore whether any of these transcripts which were not found upstream of the *ASIP* coding sequence could be located.

5.2.3 PCR investigation of the region around Exons 1a and 1b

PCR primers (Table 5.1) were designed to amplify the region around the mapping locations of *ASIP* exons 1a and 1b. The primer binding sites were based on sequence from the vicPac2 genome, and the combinations of primers used are listed in Table 5.1. Genomic DNA (50-100ng) was used as template for PCR, using 2μM of each primer, 1× RANGER reaction buffer and 0.8U RANGER DNA polymerase (Bioline) in a 10μL reaction. Amplification conditions were: 95°C for 2 min; 33 cycles of 95°C for 20s, 60°C for 30s and 72°C for 5 min; then 72°C for 10 min. Attempts to

optimise the PCR reactions which did not amplify included; extending the number of cycles to 35, decreasing the extension temperature to as low as 54°C and increasing the extension time to 10 min.

Table 5.1. Primers used to amplify the region around the mapping sites of exons 1a and 1b

Primer Name	**Sequence (5' to 3')**	**Paired with Primer:**
AT1F	GAGTTCTAAGCTCTGCACCC	AT1R, AT2R, AE2R
AT1FLR	ACAATGTCTACCAGTGAGTGG	AT1R
AT2FSR	CTGAAGGAGAGTTATCAGAGG	AT2RLR, AE2R
AT2F	GGGACAGTTCTTTGAGGTGG	AT2R, AT2RLR
AT3F	CCATAAGGCATAGTGGGTGA	AE2R
AT1R	CTCAAAGAACTGTCCCAGCC	AT1F, AT1FLR
AT2R	TAAATTATATTCCCTGAAAACGC	AT1F, AT2F
AT2RLR	CATTATCTGGCAAATAGTATGC	AT2FSR, AT2F
AE2R	AGCACAAAGGAGCTGTGACC	AT1F, AT2FSR, AT3F

5.3 Results

ASIP is most highly expressed in white animals, with moderate expression in the fawn through to light bay colour phenotypes, and little expression in the bay and black phenotypes (Figure 5.2). The gene expression of *TYRP1* negatively correlated with *ASIP* expression (Figure 5.2). *ASIP* expression decreased 10-fold between light bay to bay animals. In contrast there was a 17-fold increase in *TYRP1* expression, a threefold increase in *TYR* expression and increases in *DCT*, *PMEL17* and *MART-1* expression between the same two colour groups.

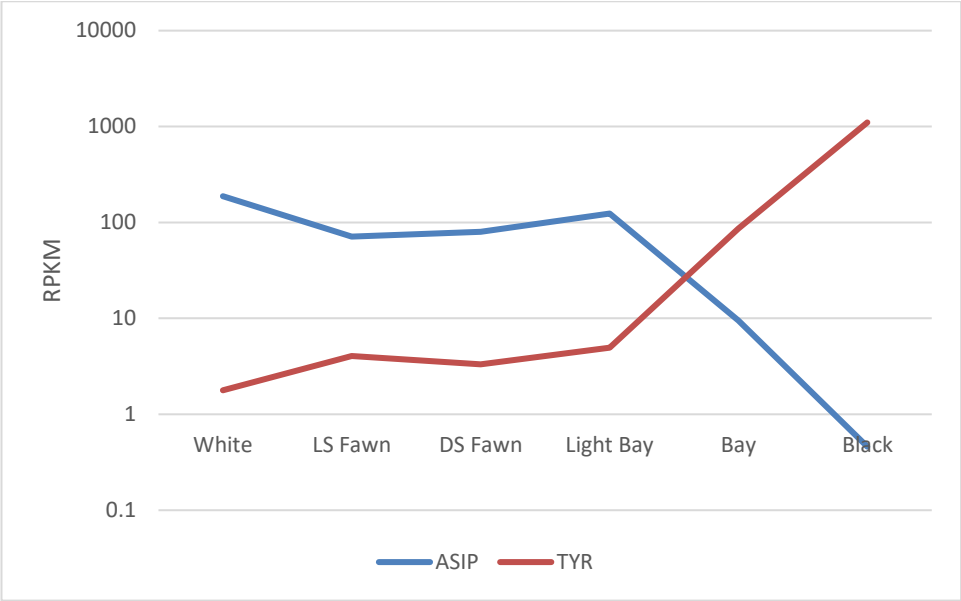

Figure 5.2. *ASIP* vs *TYRP1* gene expression levels in the different base-colours. As colour intensifies the amount of *ASIP* decreases while the rate of *TYRP1* increases, particularly in highly eumelanic black. DS – Dark Skinned, LS – Light Skinned

Four alternate 5'UTR exons were found preceding agouti exon 2, and these formed six different transcripts (Figure 5.3; Table 5.2). Exons 1a and exon 1b map to the vicPac2 genome downstream from and in the opposite orientation to *ASIP* exon 4 (Figure 5.3). Transcript 1, containing exons 1a, 2, 3 and 4, corresponds to a transcript described by Chandramohan *et al.* (2013)

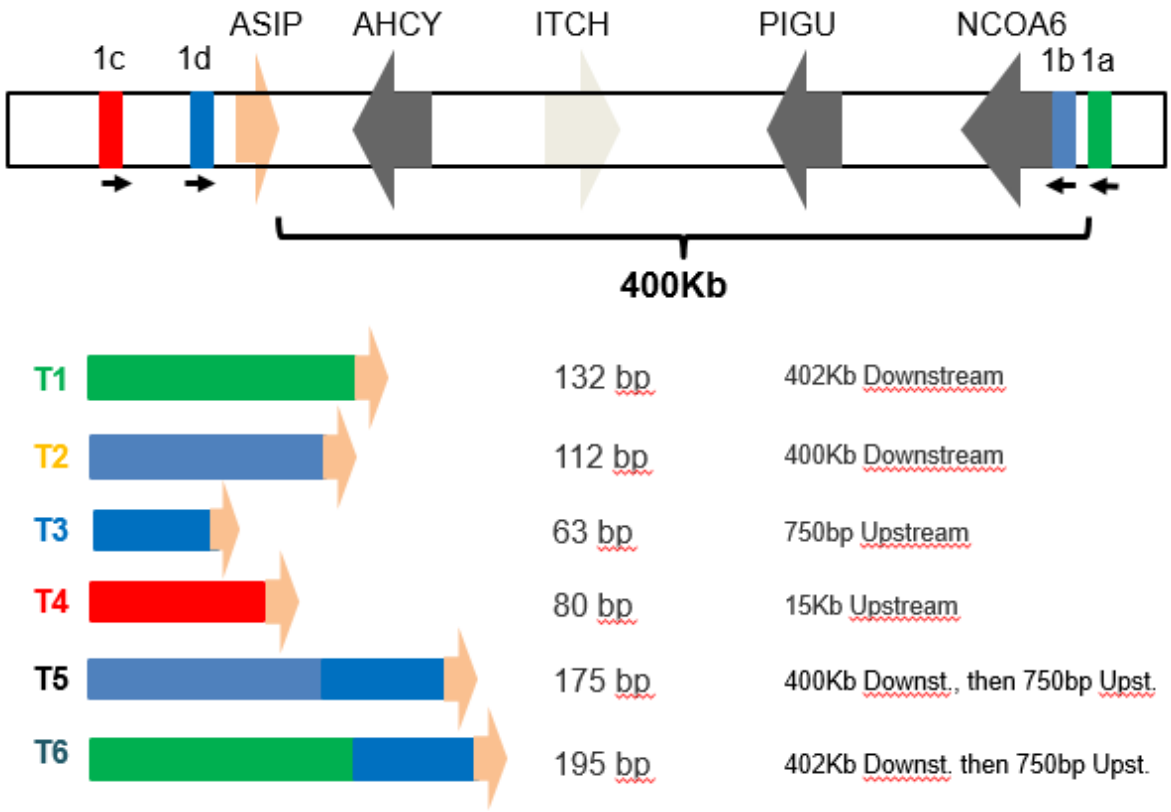

Figure 5.3. The relative positions and lengths of the 4 ASIP 5'UTR exons described, which potentially form 6 different transcripts. Adapted from Munyard and Cransberg (2016).

All six transcripts were present in the white colour groups (either in the Illumina dataset or in the white SOLiD dataset; Table 5.3). In the Illumina dataset (Chapter 3) T1 was the most prevalent transcript in all colour types whereas T2 was not found. In contrast, T2 was the most prevalent transcript in all colours in the SOLID dataset, whereas T1 was absent (Table 5.3). Exons 1a and 1b were found to either directly precede exon 2 or precede Exon 1c, which then joined *ASIP* exon 2. Transcripts 4 and 5 were found in the lighter colour groups including white, LS fawn and light bay, although at relatively low frequency (Table 5.3). It is unclear whether exon 1d exists as a single 5'UTR exon or whether it is always joined with exons 1a or 1b to form the 5' non-coding sequence of T1 or T2 respectively. None of the transcripts share homology with known 5'UTR sequences of the *ASIP* gene in other species. There appeared to be no correlation between any colour group and a particular transcript, with the majority of reads being aligned to the most commonly expressed transcript (either T1 in the Illumina dataset or T2 in the SOLiD dataset). All other transcripts were supported by too few reads to make any significant conclusions.

Table 5.2. The six alternate *ASIP* 5'UTR transcripts (denoted T1 through T6) identified. The positions listed are approximations of where the transcripts map relative to the start of exon 2 for upstream locations, or the end of exon 4 for downstream locations.

	Exon/s	Sequence Prior to the Start of Exon 2	Length	Position
T1	**1a**	CCTGAGGCCCAGCCGGGTCGGGCTGCGGGCG GCCGGGCGCCGACGGCCGGACTGACGGGCAC ACACGAGGACGGACGGGCGGACACACGGAG GGCCGCGGGCACGCACGGCCCGGGCCGGTGC TCCGAGGCCCGCCCGAGAGGGCTGGGGCCGC GCTCAG	160	~402Kb Downstream
T2	**1b**	GCCCGCCGCGCTCAGGAATCAGGGAAGATTT TGTGAACCAGTGGCACTTGAGACATGTCCTG AAGGGTAGAGGAGTTTGCCAGATGCAGAGTT GGAAGAGAAGAGCCTTCAG	112	~400Kb Downstream
T3	**1c**	CAGTAGTAAGCTGGAATGGGCCAGGATGGCA GGCCACCAAAGCAAGGAGTTTGTAGTTTTCC TCTGCACTGCATGCAAAG	80	~15Kb Upstream
T4	**1d**	GATGAAGAGAACCCTGGAAAAGCAGCCTCAG CTTTTTCCCAGACCAGAGAGAACATGCACAG	62	~750bp Upstream
T5	**1a+1d**	GCCCGCCGCGCTCAGGAATCAGGGAAGATTT TGTGAACCAGTGGCACTTGAGACATGTCCTG AAGGGTAGAGGAGTTTGCCAGATGCAGAGTT GGAAGAGAAGAGCCTTCAGGATGAAGAGAA CCCTGGAAAAGCAGCCTCAGCTTTTTCCCAG ACCAGAGAGAACATGCACAG	174	Combination of downstream and upstream
T6	**1b+1d**	CCTGAGGCCCAGCCGGGTCGGGCTGCGGGCG GCCGGGCGCCGACGGCCGGACTGACGGGCAC ACACGAGGACGGACGGGCGGACACACGGAG GGCCGCGGGCACGCACGGCCCGGGCCGGTGC TCCGAGGCCCGCCCGAGAGGGCTGGGGCCGC GCTCAGGATGAAGAGAACCCTGGAAAAGCA GCCTCAGCTTTTTCCCAGACCAGAGAGAACA TGCACAG	222	Combination of downstream and upstream

Table 5.3. Percent of Illumina and SOLiD raw read counts aligning to 5'UTR Agouti transcripts. Only reads that map to both 1a and 1c, or 1b and 1c, are included in Transcripts 5 and 6, respectively.

	Illumina Data			SOLiD Data					
Transcript	**White**	**Bay**	**Black**	**White**	**LS Fawn**	**DS Fawn**	**L Bay**	**Bay**	**Black**
T1	90.6%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
T2	0.0%	0.0%	0.0%	81.5%	57.1%	85.2%	96.9%	0.0%	100.0%
T3	7.5%	0.0%	0.0%	5.9%	0.0%	7.4%	0.0%	0.0%	0.0%
T4	0.0%	0.0%	0.0%	3.4%	0.0%	7.4%	0.0%	0.0%	0.0%
T5	0.0%	0.0%	0.0%	9.2%	42.9%	0.0%	3.1%	0.0%	0.0%
T6	1.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Exon 1a and 1b align to positions approximately 400kb downstream of Agouti Exon 3, close to, or inside the beginning of the *NCOA6* gene, respectively (Figures 5.3, 5.4).

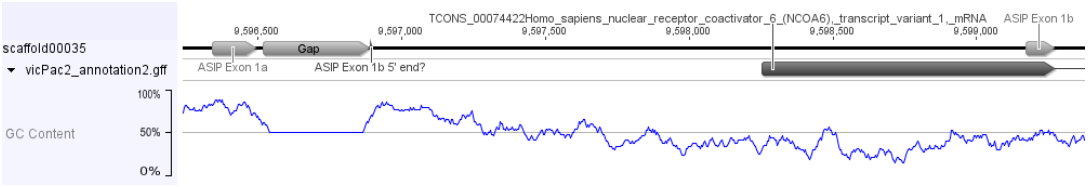

Figure 5.4. Genomic positions of the downstream transcript sequences, exon 1a and 1b. The blue line indicates GC content. A 374bp gap in the reference separates exon 1a and the potential beginning of a second position for 1b. Exon 1b is located at the 3' end of exon 1 of the NCOA6 gene.

The exon 1a sequence only matched to this one location in the vicPac2 genome, with 95% identity (Table 5.4; Figure 5.5). The 3' 132bp of this sequence was 100% identical to the sequence reported by Chandramohan *et al.* (2013, the full length sequence was 28bp longer), which was also reported to map to the same area of the genome.

Figure 5.5. Alignment of the exon 1a sequence to the vicPac2 genomic sequence. There is a decrease in similarity toward the 3' end of the alignment.

The exon 1b sequence only matched to this one location in the vicPac2 genome, and did so with 100% identity (Table 5.4). Exons 1a and 1b are located approximately 3kb apart and are separated by a 374bp gap in the reference genome (Figure 5.4). This gap is potentially caused by the high GC content of the area, which is between 75-80% in the areas around this gap (Figure 5.4).

Table 5.4. Discontiguous megablast search results searching the vicPac2 genome using the exon 1a and exon 1b sequences.

Transcript	E-Value	Sequence Length	Pairwise Identity	Query Start	Query End	Other Hits
Exon 1a	1.25e-58	159	95%	1	159	None
Exon 1b	1.26e-43	99	100%	14	112	None

Exon 1a is a homolog of the whole human *NCOA6* non-coding exon 1 sequence, while exon 1b is the 3' 99bp of the first exon found in the alpaca *NCOA6* transcript. The remaining 13bp at the 5' end of the exon 1b sequence, which doesn't align to the alpaca *NCOA6* transcript, aligns perfectly immediately following a gap in the genome (The exon 1a sequence is located 22bp upstream of this gap, Figure 5.4). This 13bp segment does not match perfectly to any other region on this particular 12.5Mb scaffold (though it does map to other locations in the genome), however due to its size and the low coverage of this section, it has not been considered to be an additional exon in this study.

The existence of both the 1a and 1b exons are supported by having many reads which pair over each side of the splice site between these exons and coding exon 2. Close to the junction of exon 1a and exon 2 in T1 there is a decrease in coverage, but this may be explained by a GC content of 84% in the 50bp before the junction (Figure 5.6). In T2 coverage remains relatively consistent across the 5'UTR (Figure 5.7).

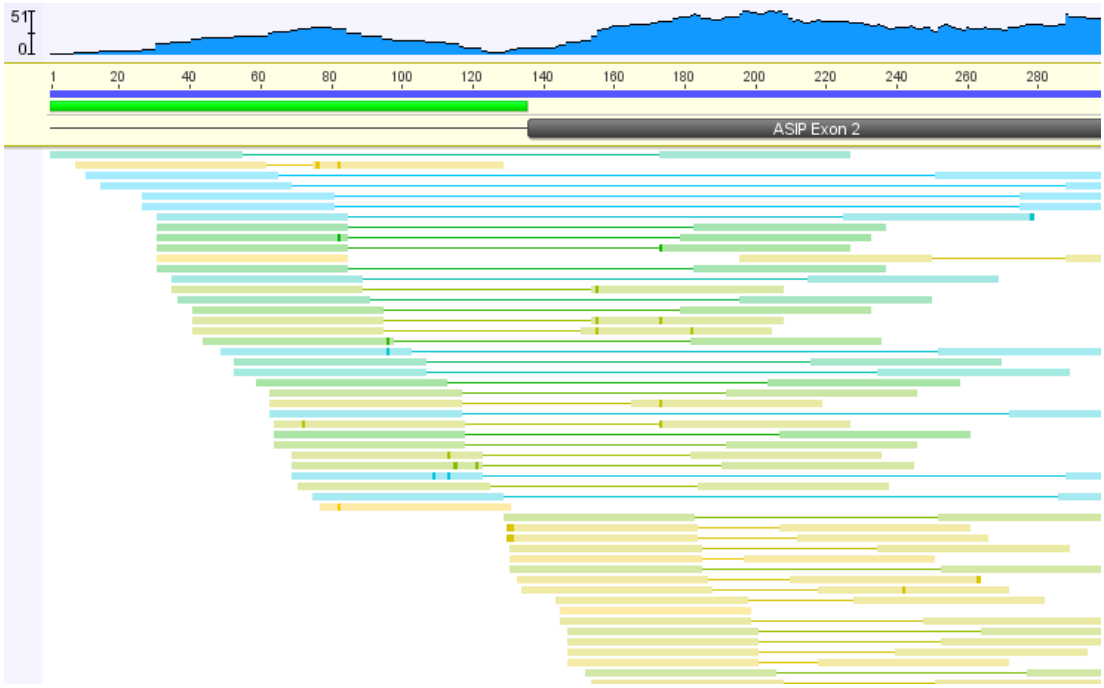

Figure 5.6. Alignment of Illumina reads from the white colour group to the T1 sequence. Reads are aligned using Geneious aligner v8.1.9 (Biomatters) with reads with more than 5% mismatches discarded. Green coloured reads indicate a match in which the paired read matches the expected distance away, yellow reads indicate a match where the paired read matches a shorter than expected distance away, blue reads indicate a match where the paired read matches further away than expected. The blue graph at the top represents the coverage. Bases in bold are indicative of a mismatch.

Figure 5.7. Alignment of SOLiD reads from the white colour group to the T2 sequence. Reads are aligned using Geneious aligner v8.1.9 (Biomatters) with reads with more than 5% mismatches discarded. Yellow coloured reads indicate a match in which the paired read matches the expected distance away, green reads indicate a match where the paired read matches a shorter than expected distance away, blue reads indicate a match where the paired read matches further away than expected. Purple Reads indicate an unpaired read. The blue graph at the top represents the coverage. Bases in bold are indicative of a mismatch.

There are a number of large gaps in the reference in the 100kb region upstream of ASIP exon 2, which includes gaps as long as 7.5Kb. Following sequencing of this region (K. Munyard, Unpublished), these gaps were filled, however exon 1a or exon 1b still showed no sequence similarity and did not map upstream of *ASIP* exon 2.

In the PCRs used to investigate the genomic structure of the region which surrounds the exons 1a and 1b, none of the primers which were designed to amplify the gap in the genomic reference between exon 1a and exon 1b were able to amplify (Figure 5.8).

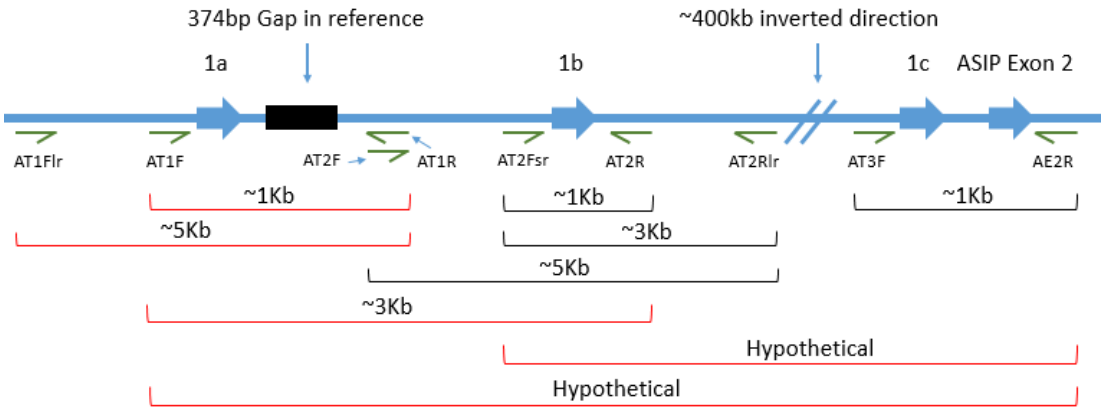

Figure 5.8. Diagrammatic representation of the primer binding sites designed to elucidate whether the structure of the reference sequence containing exons 1a and exon 1b is correct. Diagram contains a hypothesised inverted duplication and is not to scale (Genomic representation in Figure 5.3). The black brackets indicate that the PCR products matched the fragment size expected and the red brackets indicate that the target did not amplify.

5.4 Discussion

The likely cause of white in alpacas is the large increase of expression of *ASIP*. Increased *ASIP* expression has been demonstrated to have a role in causing lighter patterns, or white animals, in mice and sheep respectively (Steiner *et al.* 2007; Norris & Whan 2008; Manceau *et al.* 2011). It may play a similar role in alpacas, with white animals in particular having substantially increased levels of *ASIP* mRNA when compared to other colours. There is a strong correlation between the expression of *ASIP* and the downregulation of key colour genes, in particular *TYRP1* (Figure 5.2). Increased levels of ASIP have been demonstrated to cause downregulation of many key pigmentation genes, which results in decreases in both eumelanogenesis and pheomelanogenesis in human melanocyte cultures (Le Pape *et al.* 2008; Le Pape *et al.* 2009), a pattern which is also occurring in alpacas.

There is also the potential of *ASIP* acting to lighten other colours in alpacas. Bay and light bay animals have similar ratios of pheomelanin to eumelanin in their fibre, and their difference in colour is caused by in the amount of pheomelanin present (Cransberg *et al.* 2013). The difference in the level of *ASIP* expression is a thirteen-fold increase between light bay and bay animals respectively (Figure 5.2; Table 4.5). This increase in *ASIP* expression correlated to a sixteen-fold decrease in *TYRP1* expression, which could indicate a decreased capacity of the light bay animals to produce DHICA eumelanin. However, as all bay animals are mainly pheomelanic around the shoulder where the biopsy is taken, the increase of TYRP1 is unlikely to result in the dilution of colour observed (Cransberg et al. 2013). It is likely that the fibre colour dilution is brought about by the threefold decrease in *TYR* expression. TYR is known to function in pheomelanogenesis (Kobayashi *et al.* 1995), yet it is noted that TYR expression was not involved in the dilution of pheomelanin in French cattle (Guibert *et al.* 2004). It seems likely that the dilution observed between bay and light bay is being driven by the large increase in ASIP, but the cause of this increased expression and the pathway that causes the pheomelanin dilution is unclear.

This does not explain why the decreased *ASIP* expression and subsequent increase in *TYR* and *TYRP1* expression observed in bay, compared to the other pheomelanic

colours does not cause a eumelanic or mixed melanin phenotype. As all bay animals have eumelanic points this indicates that *MC1R* is functional, and all other colours with relatively low levels of *ASIP* (less than one tenth the level of fawn animals), are predominantly eumelanic. It seems likely that there is a threshold of *ASIP* expression below which *TYRP1* and *DCT* expression increases markedly, and it is the very high levels of expression of the *TYRP1* and *DCT* genes (which increase 13-fold and 6-fold respectively between bay and black) and subsequent proteins, which leads to the black phenotype.

In mice (Vrieling *et al.* 1994), rabbits (Fontanesi *et al.* 2010) and pigs (Drögemüller *et al.* 2006) the expression of *ASIP* varies depending on the inclusion of alternate 5'UTRs in the transcript, most of which are under the control of different promoters. In this study six alternate isoforms were discovered, and these differed only in their 5' non-coding exons (Figure 5.3; Table 5.2). One of these, here called T1, was previously described by Chandramohan *et al.* (2013) and was predicted to map 600kb downstream of ASIP exon 4 (using the vicPac1 genome). This is one of two transcripts identified in this study that aligned 400kb downstream and in the opposite orientation to the ASIP coding sequence in the VicPac2 genome, in the same position relative to other genes as the transcript described by Chandramohan *et al.* (2013) (Figure 5.3).

Transcripts T1 and T2 are not joined to ASIP exon 2 by the Tophat software, due to their orientation and distance apart from the coding sequence, they were found through alignment of the reads using Geneious (as explained in Figure 5.1). The alignments of reads from the Illumina (Figure 5.6) and SOLiD (Figure 5.7) white colour groups to T1 and T2 respectively, strongly support that these transcripts are real, with a number of fragments crossing the splice site either directly or through their respective paired reads. This is further supported through findings that these two sequences only match to this one location in the genome (as do the agouti coding exons), and only approximately 3kb apart (Figure 5.4). Coverage does drop just before the splice site in T1 between exon 1a and exon 2 but this may be due to the high GC content of the region, a factor known to decrease coverage in Illumina sequencing (Dohm *et al.* 2008). Interestingly, exon 1a has close homology to the non-coding human NCOA6 exon 1 whereas exon 1b lies at the end of the first non-

coding exon of NCOA6 identified in alpaca. Ciampolini *et al.* (2013) suggested that deregulation of promoter expression resulted in multiple 5'UTR *ASIP* start sites, which they observed in brindle boxers, some suspected to be under the control of the same promoter and some under control of different promoters. Similarly, multiple 5'UTRs regulated by the same promoter are observed in mice (Vrieling *et al.* 1994) and pig (Drögemüller *et al.* 2006).

For these distant transcripts to be processed in a way that fits the general rules of transcription some genomic recombination is required. The order of the genes that lie downstream of agouti are highly conserved in mammals (Appendix 8). It is possible that there is an inverted duplication of the agouti gene in alpacas, as suggested by Chandramohan *et al.* (2013), which is not properly represented in the vicPac1 or vicPac2 genome assemblies. The high GC content, as observed in the area around Exon 1a, has previously been demonstrated to be strongly associated with recombination events (Fullerton *et al.* 2001; Meunier & Duret 2004). Whole and partial inverted duplications of the *ASIP* gene have been reported before in sheep (Norris & Whan 2008) and mice (Chen *et al.* 1996). Alternatively, the region around the beginning of *NCOA6* could be duplicated and provide a different promoter for *ASIP*. NCOA6 mutations have been associated with pigmentation before, in melanoma studies, however these SNPs have been thought to be associated with *ASIP* variations (Liu *et al.* 2010; Sturm & Duffy 2012; Gibbs *et al.* 2015). To investigate the probability of a duplication, Repeatmasker software was used. This showed a number of repetitive elements which lie upstream of agouti, but significantly a number of erv repeats which could indicate areas for potential translocation (Appendix 9).

The gap 100 kb upstream of ASIP exon 2 was sequenced by K. Munyard (Unpublished) using long range PCR and Ion Torrent sequencing. This filled in all of the gaps that exist in the genome in this region, however exons 1a and 1b were still not found in these sequences. This could indicate either that these exons are duplicated, but further upstream, or that there is a duplication of the agouti gene, which isn't included in the vicPac2 assembly. Long-range first exons which join to *ASIP* have been demonstrated in other species, notably the *RALY* exons fusing with *ASIP* in pig, which is >150kb upstream from the *ASIP* coding sequence (Drögemüller

et al. 2006) and in Japanese quail agouti comes under the control of the *RALY* promoter – which lies over 200kb upstream of *ASIP* (Nadeau *et al.* 2008).

The last explanation for the location of these exons is that somehow during transcription, some form of long-range RNA circularisation is occurring, which is then spliced to join these regions which are 400kb downstream to the beginning of ASIP exon 2. Circularised RNAs (circRNAs) have been described in the literature (Chen 2016), but the exact mechanisms behind this phenomenon are not understood. It is unclear whether a circularised RNA of this size is possible.

It is difficult to conclude from these data why T1 and T2 are present exclusively in either the Illumina or SOLiD datasets, respectively (Table 5.3). This occurs despite the fact that all of the white animals analysed in the SOLiD dataset were also included in the previous Illumina dataset. While the Illumina and SOLiD platforms rely on different sequencing methods, it is unlikely that these could account for the differences observed, and also unlikely that any change due to sequencing technology would be able to decrease the coverage of any one isoform to 0, if we assume that both isoforms are present in both libraries, while other similar genes are expressed at similar levels to what was observed in the Illumina experiment. The type of library used does differ, with the Illumina library based on polyA captured mRNA, vs. the SOLiD library being rRNA-depleted total RNA, but this cannot logically explain the existence of this anomalous transcript. Even should, for arguments sake, one technology be malfunctioning and generating random bases – which is more likely be the SOLiD and T2, given that T1 has already been independently described, it seems too coincidental that these exons would only map to one region in the genome, approximately 3kb apart from each other.

From a biological viewpoint these samples were collected at different times from the same animals, and that could explain the differences observed. The expression of agouti has been shown to vary over time, resulting in the pattern of agouti banded hairs in mice (Bultman *et al.* 1992; Miller *et al.* 1993). Although different agouti 5' non-coding exons have previously been shown to be under the control of the same promoter and close together, there is no evidence which suggests that individual animals can express both of these isoforms at different times (Vrieling *et al.* 1994;

Drögemüller *et al.* 2006; Ciampolini *et al.* 2013), which supports the idea that perhaps collecting samples at different times was the cause of the change in isoform found.

PCR analysis of the region around exon 1a and 1b in the genome revealed a possible misassembly of this region (Figure 5.8). PCR products that were designed to cross a gap in the reference were unable to be amplified, which could indicate that the gap is longer than expected, or that the orientation of the priming sites is not as expected. Further, it could indicate a duplication where the Newbler Assembler used to determine the genomic structure was unable to determine the correct position between two points in the genome (Loman *et al.* 2012). It may also be a case of the primers not working to amplify a suitable PCR product, though PCR products of similar size were amplified using the same reaction conditions, and optimization was attempted. The other possibility is that the high GC content of the region around this gap is inhibiting the PCR (McConlogue *et al.* 1988; Hung *et al.* 1990; Pomp & Medrano 1991; Benita *et al.* 2003), as it may also have inhibited the sequencing which led to this version of the genome (Pinard *et al.* 2006).

The other exons described, exon 1c and exon 1d, share no significant homology with any of the 5' UTR exons previously described in alpacas (Chandramohan *et al.* 2013) or other species (Bultman *et al.* 1992; Drögemüller *et al.* 2006; Fontanesi *et al.* 2010; Ciampolini *et al.* 2013). It is unclear whether exon 1d exists as a standalone exon, or whether its expression is dependent on a further, upstream, non-coding exon (Either exon 1a or 1b). The relative amounts of these transcripts do not indicate that their expression is linked to a particular colour type, but, as it is a pool of samples being analysed it is possible they may determine colour in individual animals (through a combination of these exons and genotype at other colour genes). The absence of the black specific transcript described by Chandramohan *et al.* (2013) can be attributed to the relatively low number of reads which map to this locus in the black colour group, however the absence of the common transcript which was reported to be found in all three colour phenotypes which were analysed is unusual.

In conclusion, agouti has a key role in determining fibre colour in alpaca, but all of its mechanisms are not understood. It seems probable that the two downstream

5’UTR exons that are described are likely to undergo some form of chromosomal rearrangement which brings them under the control of the ASIP gene, which is not represented in the reference.

6

General Discussion

6.1 Aims and Purpose

This project aimed to elucidate causes of alpaca fibre colours through the analysis of the alpaca skin transcriptome. The majority of research into alpaca colour genetics has focussed on finding genomic polymorphisms in candidate genes. This project aimed to expand on this research by looking in combination at the expression of all genes in alpaca skin, and the melanin composition in alpaca fibre.

The general hypothesis of this thesis was that changes in the alpaca skin transcriptome would affect fibre colour. This hypothesis is supported by the results of these experiments; fibre phenotypes were objectively characterised, and RNAseq indicated that variations in the transcriptome are affecting fibre colour. The changes in expression of key colour genes, such as *ASIP*, *TYR* and *TYRP1,* indicates that alpaca pigmentation is regulated in a similar way to that in other mammalian species, and strongly suggest that differences in expression are responsible, at least in part, in the determination of alpaca fibre colour.

The three main aims of the research were all met, specifically;

Characterise the melanin present in alpaca fibre

A comparison between alpacas described as "brown" by breeders, and known eumelanic brown dogs, demonstrated a clear difference in the type of melanins present (Chapter 2). HPLC analysis confirmed that the fibre of all the "brown" alpacas sampled in this study contained predominantly pheomelanic melanin, rather than eumelanin, indicating the probable absence of the eumelanic brown phenotype in the species. These data suggest that "brown" in alpaca is predominantly caused by pheomelanin and that variations between different "brown" colours are caused by changes in pheomelanin levels, and different amounts of black eumelanin, as opposed to being brown eumelanin. Given the absence of the recessive brown allele in alpacas, alpaca breeders can make more informed mating decisions, leading to better commercial outcomes. An alteration in terminology from "brown" to bay would reflect the genotypes more accurately in alpaca.

There was no significant difference in the eumelanin: pheomelanin ratio between white and light fawn, fawn, brown, and red brown samples (Chapter 2); all of which

were predominantly pheomelanic. These data suggested a possible role for colour dilution in alpaca.

To determine whether RNAseq is a useful tool for analysing colour genes in the alpaca skin transcriptome

Analysis of the skin transcriptome with Illumina sequencing (Chapter 3) indicates that RNAseq is a tool which can be utilised to study colour genes in alpaca skin. This study enabled the first annotation of the vicPac2 genome based on biological data, which was subsequently able to be used in a more comprehensive experiment. A total of 25.4 million high quality reads were sequenced, and these were mapped to the unannotated vicPac 2.0 genome (GCA_000164845.3). Expression levels of many alpaca colour genes associated with the production of pigment in other species such as *TYR*, *MC1R* and *PMEL17* were upregulated in the black and bay skin, compared to white skin. Genes involved primarily with eumelanogenesis such as *TYRP1* and *DCT,* were highly upregulated in black skin, compared to bay and white skin, indicating similarity between regulation of alpaca pigmentation and regulation of pigmentation in other species.

Investigate the alpaca skin transcriptome.

ASIP upregulation is observed in lighter colour phenotypes which is linked to downregulation of key colour genes including *TYR*, *TYRP1*, *DCT*, *PMEL17* and *MART1*. The presence of different 5' untranslated exons observed in the *ASIP* gene could not be linked to expression or colour (Chapter 5). Sequencing libraries were created for eight different colour groups with a total of 1.4 billion reads generated from a total-RNAseq SOLiD sequencing run. The resultant reads were mapped and generated a total of 361,295 unique transcripts. GO terms analysis showed that key cellular components of skin such as keratin filaments and desmosomes were enriched in the highest expressed 100 genes. This study has improved the construction of the first alpaca skin transcriptome annotation based on biological data, by offering greater coverage which led to more complete transcripts and added additional non-coding RNA information. The role of *ASIP* as a dilution gene may be the key to understanding how to breed for white in alpaca.

6.2 The alpaca skin transcriptome in relation to fibre colour

All pheomelanic alpaca phenotypes have similar ratios of pheomelanin to eumelanin (Figure 2.8), with a continuum of intensity, so colour dilution must be occurring in alpacas. The dilution of colour is observed in many mammals, and can be caused by variations in genes such as *KIT*, *KITL*, *TYR*, *TYRP1*, *DCT*, *SLC45A2*, *SLC36A1*, *PMEL17* and *ASIP*. In horses, *KIT* mutations can result in minor colour changes such as slightly less extensive white, through to a complete white phenotype on an animal that is otherwise genetically black (Mau *et al.* 2004; Thiruvenkadan *et al.* 2008; Hauswirth *et al.* 2013). Other genes which have been shown to result in dilution phenotypes in horses include *PMEL17* (Silver locus; Locke *et al.* 2002; Brunberg *et al.* 2006) *SLC45A2* (cream, for example palomino, buckskin, perlino and cremello; Mariat *et al.* 2003), *SLC36A1* (champagne; Cook *et al.* 2008) and *MYO5A* (lavender; Brooks *et al.* 2010). In dogs, the *MLPH* gene dilutes black to blue/grey, and brown to light brown (Philipp *et al.* 2005).

There was no correlation between the expression levels of these genes and the dilution which is occurring in alpacas. However, *ASIP* expression was positively correlated with colour dilution (Table 4.5). Increased expression of *ASIP* has been shown to result in colour dilution on the ventral side of mice (Steiner *et al.* 2007; Manceau *et al.* 2011), and cause decreases in both eumelanin and pheomelanin production in human melanocyte cultures (Le Pape *et al.* 2008). In these human melanocyte cultures, increased levels of ASIP were demonstrated to downregulate the expression of many key genes of melanogenesis (Le Pape *et al.* 2009), something which is also observed in alpacas. The cause of the upregulated agouti expression in lighter colours, and in particular in white, is unclear.

Increased levels of *ASIP*, caused by a gene duplication, results in the white phenotype in sheep (Norris & Whan 2008). This duplication puts the *ASIP* gene under the control of a constitutively expressed promoter. High *ASIP* expression is also demonstrated in white alpacas, where the expression levels of *ASIP* are approximately twenty-fold higher in white skin than in bay skin, and over 200-fold higher than in black skin (in the Illumina dataset; Table 3.3), similar differences were observed in the SOLiD dataset (Table 4.5).

Confirmation of a previously reported 5'UTR exon (exon 1a), and discovery of a novel 5' UTR exon, (exon 1b), which both map 400kb downstream of *ASIP* exon 4 were found in this dataset. The presence of these exons is supported by a high number of reads crossing the splice site between these exons and exon 2 (Figures 5.6 & 5.7), and by them mapping to the same location in the genome, though 3kb apart (Table 5.4; Figure 5.4). It appears likely that there is a duplication in the genome, either in this area, or potentially whole or part of the *ASIP* gene, which brings the *ASIP* gene into a normal orientation with these 5' non-coding exons. From the available data we could not conclude that there was any correlation between the presence of these exons in *ASIP* transcripts and colour, potentially because of the low expression of *ASIP* in black and bay animals, which led to the low coverage observed in this region in these colour groups.

Increases in *ASIP* expression could also be the cause of the decrease in colour intensity between bay and light bay as this correlates with a thirteen-fold increase in *ASIP* expression (Table 4.5). The exact way in which the observed decrease in pheomelanin content (Cransberg *et al.* 2013) is regulated is unclear, however the expression of colour genes such as *TYR*, *TYRP1*, *DCT*, *PMEL17*, *MART1* and *MATP* are all downregulated between bay and light bay. Similarly the same colour genes are also decreased in white when compared against both light-skinned fawn and dark-skinned fawn, while the level of *ASIP* is over 2-fold higher in white. The upregulation of *ASIP* has a negative correlation with total melanin being produced (Figure 6.1).

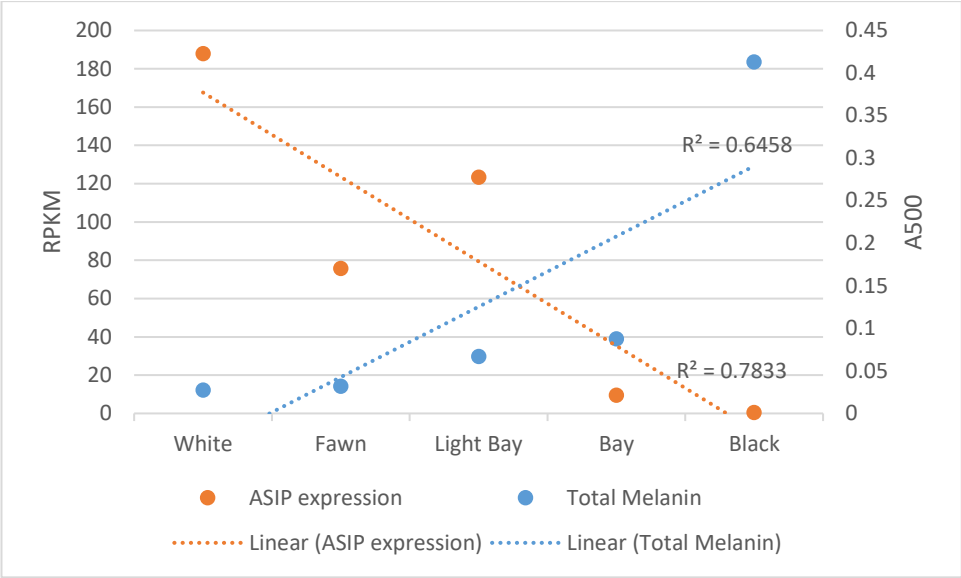

Figure 6.1 The relationship between *ASIP* expression (measured in RPKM; Chapter 4) and total melanin content (A_{500}; Chapter 2). As there was no dark-skinned fawn phenotype analysed in the melanin data the average RPKM of the light-skinned fawn and dark-skinned fawn is plotted, along with the average fawn and light fawn A_{500} value, shown here as fawn.

One potential way that *ASIP* may be regulating pigmentation is by lowering the numbers of mature melanocytes. Low levels of *MC1R* expression is observed in white skin as compared with other colours, a minimum six-fold decrease, which may imply that low numbers of melanocytes are present in white animals, or that those melanocytes which are present are not fully mature. It has been demonstrated that increased *ASIP* prevents melanocyte maturation on the ventral side of mainland and beach mice (Manceau *et al.* 2011). Low numbers, or an absence, of mature melanocytes result in white coat colour in mice (Mayer & Green 1968), pigs (Moller *et al.* 1996) and horses (Haase *et al.* 2009). Some of the largest differences in expression between white and the other colour groups was in *MART1* and *PMEL17*, which are heavily downregulated in the white group. Unexpectedly the expression levels of *MART1* and *PMEL17* were higher in light-skinned fawn than white (both approximate 6-fold increases) despite the fact that they are known to be involved in eumelanogenesis and are not required for pheomelanogenesis (Kobayashi *et al.* 1994a; Hoashi *et al.* 2005).

There are genes which are known to affect *ASIP* function which may also be impacting on phenotype. The expression level of *α-MSH* is not able to be determined

from skin because it is produced in the pituitary gland (Rajora *et al.* 1996). Other proteins, for example βDEF103, can also bind to MC1R, and the *βDEF103* variant causes dominant black in dogs (Candille *et al.* 2007). Our data showed no correlation between the expression of *βDEF103, ATRN* and *MGRN1,* all previously shown to bind or influence *ASIP* binding to *MC1R* and cause a colour change in other species (Hida *et al.* 2009), and colour phenotype in alpaca (Table 3.3; Table 4.5).

Many pigmentation genes are transcriptionally regulated by MITF, which is also the master regulator of melanocyte development (Levy *et al.* 2006; Vachtenheim & Borovanský 2010). MITF has been demonstrated to regulate expression of colour genes including *TYR, TYRP1, DCT, MC1R, PMEL17, MART1* and *RAB27A* (Vachtenheim & Borovanský 2010). MITF has also been demonstrated to upregulate *SLC45A2* expression. This appears to be an indirect effect, despite *SLC45A2* having several potential MITF binding motifs (Du & Fisher 2002). Despite being under the control of the same transcription factor, the expression of these genes falls into three groups in the Illumina data: *TYR, MC1R* and *SLC45A2* all have very low expression in white skin, moderate in bay, and approximately double the bay expression levels in black; *TYRP1* and *DCT* have relatively low expression in white and bay and much higher expression in black; whereas *PMEL17* and *MART1* have low expression in white and high expression in bay and black (Table 3.3). These results are similar in the SOLiD dataset, though with different fold-changes observed (Table 4.5). However, there is no significant variation between *MITF* gene expression between the different colours (Table 4.5). No polymorphisms were observed in the reads that mapped to the *MITF* transcript (Data not shown).

Increased *ASIP* expression is associated with lower expression levels of the tyrosinase family of genes; *TYR, TYRP1* and *DCT*, which encode the key enzymes of melanogenesis (Lamoreux *et al.* 2001). This is observed in both the Illumina and SOLiD experiments, with low expression of this gene family in white, an increase in expression in bay and highest expression in black while ASIP expression has the opposite pattern (Table 3.3, Table 4.5). It has been demonstrated that increase levels of TYR and TYRP1 can have a role in switching to eumelanogenesis from pheomelanogenesis (Ito *et al.* 2000; Lamoreux *et al.* 2001). *TYR* expression in bay

skin is more similar to the levels observed in black skin (less than two-fold increase) than either *TYRP1* or *DCT*, which both display large fold expression increases (6-fold and 13-fold in the SOLiD dataset) in black compared to bay in both datasets. Given that bay animals are predominantly pheomelanic in the area which was sampled (Cransberg *et al.* 2013 & Chapter 2) the large increases of expression observed are expected, given that *TYRP1* and *DCT* are not essential for pheomelanin production in mice (Kobayashi *et al.* 1995). This may also be indicative of a more similar pigmentation pathway to mice rather than humans. In mice *TYRP1* has been shown to be primarily responsible for the oxidation of DHICA in the formation of eumelanin, a function which is absent in humans (Kobayashi *et al.* 1994b). Given the high DHICA to DHI ratios seen in alpaca eumelanin (Cransberg *et al.* 2013 & Chapter 2) it is reasonable to expect that much higher levels of *TYRP1* would be needed to create the large amounts of DHICA present in black alpaca skin (Figure 2.11). *TYRP1* expression is much higher in black than white skin, which is in contrast to a skin transcriptome study in sheep where it was reported that *TYR* had a larger difference in expression than *TYRP1* between white and black animals (Fan *et al.* 2013). An alternate function of *TYRP1*, other than DHICA oxidation, is stabilising TYR, which is seen in mice but not in humans (Boissy *et al.* 1998; Kobayashi *et al.* 1998). The high expression of *TYRP1* observed in black alpacas could possibly lead to large amounts of *TYRP1* being available to stabilise TYR, making TYR more effective in eumelanin synthesis.

The same expression pattern is observed in *MATP* and *TYR* across the different skin colours, which is also observed in mice (Baxter & Pavan 2002). Though the function of *SLC45A2* in pigmentation is not yet fully understood, the cause of the similar expression levels of *TYR* and *SLC45A2* observed between different coloured alpaca skin could potentially be caused by *MATP* increasing the levels of *TYR* expression through increasing the pH in melanocytes (Newton *et al.* 2001; Dooley *et al.* 2013). It has also been reported that the trafficking and processing of TYR is disrupted in underwhite mice, a phenotype caused by a less effective MATP protein, which results in less TYR being released into immature melanosomes, which results in a hypo-pigmented phenotype (Costin *et al.* 2003). The third and most likely reason for the similar pattern of expression between *TYR* and *MATP* in different coloured alpaca skin is a common regulatory signal, either through the action of the same

transcription factor in each gene, MITF (Baxter & Pavan 2002; Yamaguchi *et al.* 2007), or by the interactions between MC1R and ASIP.

6.3 Limitations

The biggest limitations of this work came from the design of the SOLiD experiment (Chapter 4). While the choice of total-RNA included previously undocumented non-coding RNAs, which has been useful in identifying non-coding RNAs, such as snoRNAs, which could be influential on colour, this also included nascent RNAs into the dataset. These nascent RNAs are expected to be responsible for some 3' read-through, as well as creating some isoforms which may be incorrect because of incomplete splicing of introns. The contribution of these processes is difficult to quantitate as exactly what is expected in alpaca is unknown, and consensuses markers such as polyadenylation sites, which may be used to trim these sequences, could bias the data further. It was necessary to include these data in the genome, as the Illumina reads alone (Chapter 3) provided insufficient coverage to achieve full length transcript sequences for many genes, and the extra coverage provided with the SOLiD reads markedly improved transcript completeness.

The choice of the SOLiD platform was made because it was considered to be the most accurate technology at the time the experiment was conducted. The subsequent collapse of the SOLiD brand, and the fact that the proprietary software is extremely difficult to install meant that the data could not be re-analysed after the release of the vicPac2 genome. As demonstrated in chapter 4, the use of SOLiD reads outside of their proprietary software was challenging, and severely limits the tools available to analyse the data. Furthermore, it is impossible to create a *de novo* transcriptome assembly with SOLiD data, which could have been used to identify poly-A tailed mRNA in the dataset.

6.4 Future Directions

The most intriguing aspect of these data is the presence of 5'UTR exons of the *ASIP* gene 400kb downstream of *ASIP* exon 4. The upcoming release of the vicPac3 genome, which will be based on PacBio long reads and high coverage Illumina reads

(Mark Richardson, personal communication) may reveal how this occurs. Should the new genome assembly fail to provide a solution, it would be useful to apply the Hi-C technique (Belton *et al.* 2012), which looks for DNA-DNA interactions in a similar process to that used in CHiP-Seq (Schmidt *et al.* 2009), to determine if these interactions are real. ATAC-seq could also be used to look for regulatory regions that are active in different colours (Buenrostro *et al.* 2015). It will also be interesting to determine whether the change between T1 and T2 is in-fact a seasonal change.

This dataset contains a lot of information which is yet to be fully analysed, such as looking at alternative isoform abundance, miRNAs, and a complete analysis of the roan and silvergrey patterns, both of which expected to be caused by incomplete dominant alleles.

6.5 Conclusions

Upregulation of the *ASIP* gene has an important role in determining the white colour in alpacas, and likely also plays a role in colour dilution. Increased *ASIP* expression, which could result from a whole or partial gene duplication, potentially causes a reduction in the number of mature melanocytes, and causes the downregulation of many key colour genes including *TYR*, *TYRP1*, *DCT*, *PMEL17*, *MART1* and *MATP*, which results in lower melanin production.

Chemical melanin analysis revealed that animals being described as brown were actually closer to the bay colour in horses as opposed to eumelanic brown (Cransberg *et al.* 2013). These studies, in conjunction with those by Feeley *et al.* (2016) have demonstrated that eumelanic brown either does not exist in Australian alpacas, or exists only in a small percentage of the population.

RNAseq analysis of alpaca skin in eight common phenotypes has vastly increased the knowledge of alpaca colour genetics. The first alpaca skin transcriptome based on biological data was created as part of this study. This was also the first study to investigate the likely roles and functions of key genes and pathways of pigmentation in alpaca. The expression of key colour genes indicates that these genes have similar roles in alpacas to other species, particularly mice.

References

Abel J.J. & Davis W.S. (1896) On the pigment of the Negro's skin and hair. *The Journal of Experimental Medicine* **1**, 361-400.

Aksan I. & Goding C. (1998) Targeting the microphthalmia basic helix-loop-helix–leucine zipper transcription factor to a subset of E-box elements in vitro and in vivo. *Molecular and Cellular Biology* **18**, 6930-8.

Albinus B.S. (1737) *Icones ossium foetus humani*. apud Joh. et Herm. Verbeek.

Aliev G., Rachkovsky M., Ito S., Wakamatsu K. & Ivanov A. (1990) Pigment Types in Selected Color Genotypes of Asiatic Sheep. *Pigment Cell Research* **3**, 177-80.

Ameur A., Zaghlool A., Halvardson J., Wetterbom A., Gyllensten U., Cavelier L. & Feuk L. (2011) Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nature Structural & Molecular Biology* **18**, 1435-40.

Ancans J., Tobin D.J., Hoogduijn M.J., Smit N.P., Wakamatsu K. & Thody A.J. (2001) Melanosomal pH Controls Rate of Melanogenesis, Eumelanin/Phaeomelanin Ratio and Melanosome Maturation in Melanocytes and Melanoma Cells. *Experimental Cell Research* **268**, 26-35.

Ansorge W.J. (2009) Next-generation DNA sequencing techniques. *New Biotechnology* **25**, 195-203.

Aoki H., Motohashi T., Yoshimura N., Yamazaki H., Yamane T., Panthier J.J. & Kunisada T. (2005) Cooperative and indispensable roles of endothelin 3 and KIT signalings in melanocyte development. *Developmental Dynamics* **233**, 407-17.

Aydin I.T., Hummler E., Smit N.P.M. & Beermann F. (2012) Coat color dilution in mice because of inactivation of the melanoma antigen MART-1. *Pigment Cell and Melanoma Research* **25**, 37-46.

Aylan-Parker J. & McGregor B.A. (2002) Optimising sampling techniques and estimating sampling variance of fleece quality attributes in alpacas. *Small Ruminant Research* **44**, 53-64.

Bagher P., Jiao J., Smith C.O., Cota C.D. & Gunn T.M. (2006) Characterization of Mahogunin Ring Finger-1 expression in mice. *Pigment Cell Research* **19**, 635-43.

Barsh G.S. (2006) Regulation of pigment type switching by agouti, melanocortin signaling, attractin, and mahoganoid. *The Pigmentary System: Physiology and Pathophysiology, Second Edition*, 395-409.

Baxter L.L. & Pavan W.J. (2002) The oculocutaneous albinism type IV gene Matp is a new marker of pigment cell precursors during mouse embryonic development. *Mechanisms of Development* **116**, 209-12.

Belton J.-M., McCord R.P., Gibcus J.H., Naumova N., Zhan Y. & Dekker J. (2012) Hi–C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268-76.

Benita Y., Oosting R.S., Lok M.C., Wise M.J. & Humphery-Smith I. (2003) Regionalized GC content of template DNA as a predictor of PCR success. *Nucleic Acids Research* **31**, e99.

Bennett D. (1991) Colour genes, oncogenes and melanocyte differentiation. *Journal of Cell Science* **98**, 135-9.

Bennett D.C. & Lamoreux M.L. (2003) The Color Loci of Mice - A Genetic Century. *Pigment Cell Research* **16**, 333-44.

Berson J.F., Harper D.C., Tenza D., Raposo G. & Marks M.S. (2001) Pmel17 Initiates Premelanosome Morphogenesis within Multivesicular Bodies. *Molecular Biology of the Cell* **12**, 3451-64.

Blankenberg D., Gordon A., Von Kuster G., Coraor N., Taylor J., Nekrutenko A. & Team G. (2010) Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783-5.

Boissy R.E., Sakai C., Zhao H., Kobayashi T. & Hearing V.J. (1998) Human tyrosinase related protein-1 (TRP-1) does not function as a DHICA oxidase activity in contrast to murine TRP-1. *Experimental Dermatology* **7**, 198-204.

Brooks S.A., Gabreski N., Miller D., Brisbin A., Brown H.E., Streeter C., Mezey J., Cook D. & Antczak D.F. (2010) Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. *PLoS Genetics* **6**, e1000909.

Brunberg E., Andersson L., Cothran G., Sandberg K., Mikko S. & Lindgren G. (2006) A missense mutation in PMEL17 is associated with the Silver coat color in the horse. *BMC Genetics* **7**, 46.

Buenrostro J.D., Wu B., Chang H.Y. & Greenleaf W.J. (2015) ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology*, 21.9. 1-.9. 9.

Bultman S.J., Michaud E.J. & Woychik R.P. (1992) Molecular characterization of the mouse agouti locus. *Cell* **71**, 1195-204.

Busca R. & Ballotti R. (2000) Cyclic AMP a Key Messenger in the Regulation of Skin Pigmentation. *Pigment Cell Research* **13**, 60-9.

Butterfield L.H., Stoll T.C., Lau R. & Economou J.S. (1997) Cloning and analysis of MART-1/Melan-A human melanoma antigen promoter regions. *Gene* **191**, 129-34.

Camacho-Hübner A., Richard C. & Beermann F. (2002) Genomic structure and evolutionary conservation of the tyrosinase gene family from Fugu. *Gene* **285**, 59-68.

Camp E., Badhwar P., Mann G.J. & Lardelli M. (2003) Expression Analysis of a *Tyrosinase* Promoter Sequence in Zebrafish. *Pigment Cell Research* **16**, 117-26.

Candille S.I., Kaelin C.B., Cattanach B.M., Yu B., Thompson D.A., Nix M.A., Kerns J.A., Schmutz S.M., Millhauser G.L. & Barsh G.S. (2007) A β-Defensin Mutation Causes Black Coat Color in Domestic Dogs. *Science* **318**, 1418-23.

Cecchi T., Cozzali C., Passamonti P., Ceccarelli P., Pucciarelli F., Gargiulo A.M., Frank E.N. & Renieri C. (2004) Melanins and Melanosomes From Llama (Lama glama L.). *Pigment Cell Research* **17**, 307-11.

Chakraborty A.K., Platt J.T., Kim K.K., Kwon B.S., Bennet D.C. & Pawelek J.M. (1996) Polymerization of 5,6-dihydroxyindole-2-carboxylic acid to melanin by the pmel 17 silver locus protein. *European Journal of Biochemistry* **236**, 180-8.

Chandramohan B., Renieri C., Manna V.L. & Terza A.L. (2013) The alpaca *agouti* gene: Genomic locus, transcripts and causative mutations of eumelanic and pheomelanic coat color. *Gene* **521**, 303-10.

Chen L.-L. (2016) The biogenesis and emerging roles of circular RNAs. *Nature Reviews Molecular Cell Biology* **17**, 205-11.

Chen Y., Duhl D.M. & Barsh G.S. (1996) Opposite orientations of an inverted duplication and allelic variation at the mouse agouti locus. *Genetics* **144**, 265-77.

Chiaverini C., Beuret L., Flori E., Busca R., Abbe P., Bille K., Bahadoran P., Ortonne J.-P., Bertolotto C. & Ballotti R. (2008) Microphthalmia-associated Transcription Factor Regulates RAB27A Gene Expression and Controls Melanosome Transport. *Journal of Biological Chemistry* **283**, 12635-42.

Ciampolini R., Cecchi F., Spaterna A., Bramante A., Bardet S.M. & Oulmouden A. (2013) Characterization of different 5′-untranslated exons of the ASIP gene in black-and-tan Doberman Pinscher and brindle Boxer dogs. *Animal Genetics* **44**, 114-7.

Cicero A.L., Delevoye C., Gilles-Marsens F., Loew D., Dingli F., Guéré C., André N., Vié K., van Niel G. & Raposo G. (2015) Exosomes released by keratinocytes modulate melanocyte pigmentation. *Nature Communications* **6**.

Cook D., Brooks S., Bellone R. & Bailey E. (2008) Missense mutation in exon 2 of SLC36A1 responsible for champagne dilution in horses. *PLoS Genetics* **4**, e1000195.

Cooper C.D. & Raible D.W. (2009) Mechanisms for reaching the differentiated state: Insights from neural crest-derived melanocytes. *Seminars in Cell & Developmental Biology* **20**, 105-10.

Cortese K., Giordano F., Surace E.M., Venturi C., Ballabio A., Tacchetti C. & Marigo V. (2005) The ocular albinism type 1 (OA1) gene controls melanosome maturation and size. *Investigative Ophthalmology and Visual Science* **46**, 4358-64.

Costa V., Aprile M., Esposito R. & Ciccodicola A. (2013) RNA-Seq and human complex diseases: recent accomplishments and future perspectives. *European Journal of Human Genetics* **21**, 134-42.

Costin G.-E. & Hearing V.J. (2007) Human skin pigmentation: melanocytes modulate skin color in response to stress. *The FASEB Journal* **21**, 976-94.

Costin G.-E., Valencia J.C., Vieira W.D., Lamoreux M.L. & Hearing V.J. (2003) Tyrosinase processing and intracellular trafficking is disrupted in mouse primary melanocytes carrying the underwhite (uw) mutation. A model for oculocutaneous albinism (OCA) type 4. *Journal of Cell Science* **116**, 3203-12.

Costin G.-E., Valencia Julio C., Wakamatsu K., Ito S., Solano F., Milac Adina L., Vieira Wilfred D., Yamaguchi Y., Rouzaud F., Petrescu A.-J., Lamoreux M L. & Hearing Vincent J. (2005) Mutations in dopachrome tautomerase (Dct) affect eumelanin/pheomelanin synthesis, but do not affect intracellular trafficking of the mutant protein. *Biochemical Journal* **391**, 249-59.

Cransberg R. & Munyard K.A. (2011) Polymorphisms detected in the tyrosinase and matp (slc45a2) genes did not explain coat colour dilution in a sample of Alpaca (Vicugna pacos). *Small Ruminant Research* **95**, 92-6.

Cransberg R., Wakamatsu K. & Munyard K. (2013) Melanin characterisation suggests that the “brown” phenotype in alpaca (Vicugna pacos) is predominantly pheomelanic. *Small Ruminant Research* **114**, 240-6.

Cuénot L. (1902) Notes et revues. *Arch. Zool. Exp. Gen*.

Davison I. (2008) The Coming of A.G.E.: Breeding a Better Alpaca. In: *World Alpaca Conference*, Sydney.

Deng W.D., Xi D.M., Gou X., Yang S.L., Shi X.W. & Mao H.M. (2007) Pigmentation in Black Boned sheep (Ovis aries): association with polymorphism of the tyrosinase gene. *Molecular Biology Reports* **35**, 379-85.

Dohm J.C., Lottaz C., Borodina T. & Himmelbauer H. (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research* **36**, e105.

Dooley C.M., Schwarz H., Mueller K.P., Mongera A., Konantz M., Neuhauss S.C., Nüsslein-Volhard C. & Geisler R. (2013) Slc45a2 and V-ATPase are regulators of

melanosomal pH homeostasis in zebrafish, providing a mechanism for human pigment evolution and disease. *Pigment Cell & Melanoma Research* **26**, 205-17.

Drögemüller C., Giese A., Martins-Wess F., Wiedemann S., Andersson L., Brenig B., Fries R. & Leeb T. (2006) The mutation causing the black-and-tan pigmentation phenotype of Mangalitza pigs maps to the porcine ASIP locus but does not affect its coding sequence. *Mammalian Genome* **17**, 58-66.

Du J. & Fisher D.E. (2002) Identification of Aim-1 as the *underwhite* Mouse Mutant and Its Transcriptional Regulation by MITF. *Journal of Biological Chemistry* **277**, 402-6.

Du J., Miller A.J., Widlund H.R., Horstmann M.A., Ramaswamy S. & Fisher D.E. (2003) MLANA/MART1 and SILV/PMEL17/GP100 Are Transcriptionally Regulated by MITF in Melanocytes and Melanoma. *The American Journal of Pathology* **163**, 333-43.

Ender C., Krek A., Friedländer M.R., Beitzinger M., Weinmann L., Chen W., Pfeffer S., Rajewsky N. & Meister G. (2008) A human snoRNA with microRNA-like functions. *Molecular Cell* **32**, 519-28.

Fan R., Gang Y. & Changsheng D. (2010) Study of Hair Melanins in Various Hair Color Alpaca (Lama Pacos). *Asian-Australasian Journal of Animal Science* **23**, 444-9.

Fan R., Xie J., Bai J., Wang H., Tian X., Bai R., Jia X., Yang L., Song Y., Herrid M., Gao W., He X., Yao J., Smith G. & Dong C. (2013) Skin transcriptome profiles associated with coat color in sheep. *BMC Genomics* **14**, 389.

Feeley N.L., Bottomley S. & Munyard K.A. (2011) Three novel mutations in ASIP associated with black fibre in alpacas (Vicugna pacos). *The Journal of Agricultural Science* **149**, 529-38.

Feeley N.L., Bottomley S. & Munyard K.A. (2016) Novel mutations in Vicugna pacos (alpaca) Tyrp1 are not correlated with brown fibre colour phenotypes. *Small Ruminant Research* **143**, 29-34.

Feeley N.L. & Munyard K.A. (2009) Characterisation of the melanocortin-1 receptor gene in alpaca and identification of possible markers associated with phenotypic variations in colour. *Animal Production Science* **49**, 675-81.

Fitzpatrick T.B., Miyamoto M. & Ishikawa K. (1967) The evolution of concepts of melanin biology. *Archives of Dermatology* **96**, 305-23.

Fontanesi L., Forestier L., Allain D., Scotti E., Beretti F., Deretz-Picoulet S., Pecchioli E., Vernesi C., Robinson T.J. & Malaney J.L. (2010) Characterization of the rabbit agouti signaling protein (ASIP) gene: transcripts and phylogenetic analyses and identification of the causative mutation of the nonagouti black coat colour. *Genomics* **95**, 166-75.

Frank E.N., Hick M.V.H., Gauna C.D., Lamas H.E., Renieri C. & Antonini M. (2006) Phenotypic and genetic description of fibre traits in South American domestic camelids (llamas and alpacas). *Small Ruminant Research* **61**, 113-29.

Freedman M.L., Monteiro A.N., Gayther S.A., Coetzee G.A., Risch A., Plass C., Casey G., De Biasi M., Carlson C. & Duggan D. (2011) Principles for the post-GWAS functional characterization of cancer risk loci. *Nature Genetics* **43**, 513-8.

Fullerton S.M., Carvalho A.B. & Clark A.G. (2001) Local rates of recombination are positively correlated with GC content in the human genome. *Molecular Biology and Evolution* **18**, 1139-42.

Furumura M., Sakai C., Abdel-Malek Z., Barsh G.S. & Hearing V.J. (1996) The Interaction of Agouti Signal Protein and Melanocyte Stimulating Hormone to Regulate Melanin Formation in Mammals. *Pigment Cell Research* **9**, 191-203.

Gibbs D.C., Orlow I., Kanetsky P.A., Luo L., Kricker A., Armstrong B.K., Anton-Culver H., Gruber S.B., Marrett L.D. & Gallagher R.P. (2015) Inherited genetic variants associated with occurrence of multiple primary melanoma. *Cancer Epidemiology Biomarkers & Prevention* **24**, 992-7.

Giordano F., Bonetti C., Surace E.M., Marigo V. & Raposo G. (2009) The ocular albinism type 1 (OA1) G-protein-coupled receptor functions with MART-1 at early stages of melanogenesis to control melanosome identity and composition. *Human Molecular Genetics* **18**, 4530-45.

Girardot M., Martin J., Guibert S., Leveziel H., Julien R. & Oulmouden A. (2005) Widespread expression of the bovine Agouti gene results from at least three alternative promoters. *Pigment Cell Research* **18**, 34-41.

Gordon A. & Hannon G. (2010) Fastx-toolkit. *FASTQ/A short-reads pre-processing tools (unpublished) http://hannonlab. cshl. edu/fastx_toolkit*.

Görnitz K. (1923) Versuch einer Klassifikation der häufigsten Federfärbungen. *Journal of Ornithology* **71**, 127-31.

Graf J., Voisey J., Hughes I. & Daal A.v. (2007) Promoter polymorphisms in the *MATP* (*SLC45A2*) gene are associated with normal human skin color variation. *Human Mutation* **28**, 710-7.

Guibert S., Girardot M., Leveziel H., Julien R. & Oulmouden A. (2004) Pheomelanin Coat Colour Dilution in French Cattle Breeds is not Correlated with the TYR, TYRP1 and DCT Transcription Levels. *Pigment Cell Research* **17**, 337-45.

Haase B., Brooks S., Tozaki T., Burger D., Poncet P.A., Rieder S., Hasegawa T., Penedo C. & Leeb T. (2009) Seven novel KIT mutations in horses with white coat colour phenotypes. *Animal Genetics* **40**, 623-9.

Harper D.C., Theos A.C., Herman K.E., Tenza D., Raposo G. & Marks M.S. (2008) Premelanosome Amyloid-like Fibrils Are Composed of Only Golgi-processed Forms of Pmel17 That Have Been Proteolytically Processed in Endosomes. *Journal of Biological Chemistry* **283**, 2307-22.

Hauswirth R., Jude R., Haase B., Bellone R.R., Archer S., Holl H., Brooks S.A., Tozaki T., Penedo M.C.T. & Rieder S. (2013) Novel variants in the KIT and PAX3 genes in horses with white-spotted coat colour phenotypes. *Animal Genetics* **44**, 763-5.

He L., Eldridge A.G., Jackson P.K., Gunn T.M. & Barsh G.S. (2003) Accessory proteins for melanocortin signaling - Attractin and mahogunin. *Melanocortin System* **994**, 288-98.

Henle J. (1837) *Symbolae ad anatomiam villorum intestinalium, imprimis eorum epithelii et vasorum lacteorum*. prostat apud Aug. Hirschwald.

Hida T., Wakamatsu K., Sviderskaya E.V., Donkin A.J., Montoliu L., Lynn Lamoreux M., Yu B., Millhauser G.L., Ito S. & Barsh G.S. (2009) Agouti protein, mahogunin, and attractin in pheomelanogenesis and melanoblast-like alteration of melanocytes: a cAMP-independent pathway. *Pigment Cell and Melanoma Research* **22**, 623-34.

Hirobe T., Wakamatsu K., Ito S., Kawa Y., Soma Y. & Mizoguchi M. (2006) The slaty mutation affects eumelanin and pheomelanin synthesis in mouse melanocytes. *European Journal of Cell Biology* **85**, 537-49.

Hoashi T., Watabe H., Muller J., Yamaguchi Y., Vieira W.D. & Hearing V.J. (2005) MART-1 Is Required for the Function of the Melanosomal Matrix Protein PMEL17/GP100 and the Maturation of Melanosomes. *Journal of Biological Chemistry* **280**, 14006-16.

Hoekstra H.E. (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* **97**, 222-34.

Hou L., Panthier J.J. & Arnheiter H. (2000) Signaling and transcriptional regulation in the neural crest-derived melanocyte lineage: interactions between KIT and MITF. *Development* **127**, 5379-89.

Hrdlickova R., Toloue M. & Tian B. (2016) RNA-Seq methods for transcriptome analysis. *Wiley Interdisciplinary Reviews: RNA*.

Huang D.W., Sherman B.T. & Lempicki R.A. (2009a) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**, 1-13.

Huang D.W., Sherman B.T. & Lempicki R.A. (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57.

Hume A.N., Ushakov D.S., Tarafder A.K., Ferenczi M.A. & Seabra M.C. (2007) Rab27a and MyoVa are the primary Mlph interactors regulating melanosome transport in melanocytes. *Journal of Cell Science* **120**, 3111-22.

Hung T., Mak K. & Fong K. (1990) A specificity enhancer for polymerase chain reaction. *Nucleic Acids Research* **18**, 4953.

Hurst C.C. (1905) Experimental Studies on Heredity in Rabbits. *Journal of the Linnean Society of London, Zoology* **29**, 283-324.

Hurst C.C. (1906) On the Inheritance of Coat Colour in Horses. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* **77**, 388-94.

Hustad C.M., Perry W.L., Siracusa L.D., Rasberry C., Cobb L., Cattanach B.M., Kovatch R., Copeland N.G. & Jenkins N.A. (1995) Molecular genetic characterization of six recessive viable alleles of the mouse agouti locus. *Genetics* **140**, 255-65.

Incerti B., Cortese K., Pizzigoni A., Surace E.M., Varani S., Coppola M., Jeffery G., Seeliger M., Jaissle G., Bennett D.C., Marigo V., Schiaffino M.V., Tacchetti C. & Ballabio A. (2000) Oa1 knock-out: new insights on the pathogenesis of ocular albinism type 1. *Human Molecular Genetics* **9**, 2781-8.

Ito S. (2003) A Chemist's View of Melanogenesis. *Pigment Cell Research* **16**, 230-6.

Ito S. & Wakamatsu K. (1998) Chemical Degradation of Melanins: Application to Identification of Dopamine-melanin. *Pigment Cell Research* **11**, 120-6.

Ito S. & Wakamatsu K. (2003) Quantitative Analysis of Eumelanin and Pheomelanin in Humans, Mice, and Other Animals: a Comparative Review. *Pigment Cell Research* **16**, 523-31.

Ito S. & Wakamatsu K. (2008) Chemistry of Mixed Melanogenesis—Pivotal Roles of Dopaquinone†. *Photochemistry and Photobiology* **84**, 582-92.

Ito S., Wakamatsu K. & Ozeki H. (2000) Chemical analysis of melanins and its application to the study of the regulation of melanogenesis. *Pigment Cell Research* **13**, 103-9.

Jackson I., Chambers D., Tsukamoto K., Copeland N., Gilbert D., Jenkins N. & Hearing V. (1992) A second tyrosinase-related protein, TRP-2, maps to and is mutated at the mouse slaty locus. *The EMBO journal* **11**, 527.

Jimbow K., Fitzpatrick T. & Wick M. (1991) Biochemistry and physiology of melanin pigmentation. *Physiology, Biochemistry, and Molecular Biology of the Skin* **2**, 873-909.

Jimbow K., Hua C., Gomez P.F., Hirosaki K., Shinoda K., Salopek T.G., Matsusaka H., Jin H.-Y. & Yamashita T. (2000) Intracellular Vesicular Trafficking of Tyrosinase Gene Family Protein in Eu- and Pheomelanosome Biogenesis. *Pigment Cell Research* **13**, 110-7.

Kadwell M., Fernandez M., Stanley H.F., Baldi R., Wheeler J.C., Rosadio R. & Bruford M.W. (2001) Genetic analysis reveals the wild ancestors of the llama and the alpaca. *Proceedings. Biological sciences / The Royal Society* **268**, 2575-84.

Kent W.J. (2002) BLAT—The BLAST-Like Alignment Tool. *Genome Research* **12**, 656-64.

Kijas J.M.H., Wales R., Tornsten A., Chardon P., Moller M. & Andersson L. (1998) Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. *Genetics* **150**, 1177-85.

Kishore S. & Stamm S. (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* **311**, 230-2.

Kobayashi T., Imokawa G., Bennett D.C. & Hearing V.J. (1998) Tyrosinase Stabilization by Tyrp1 (the brown Locus Protein). *Journal of Biological Chemistry* **273**, 31801-5.

Kobayashi T., Urabe K., Orlow S.J., Higashi K., Imokawa G., Kwon B.S., Potterf B. & Hearing V.J. (1994a) The pmel-17 silver locus protein - characterization and investigation of its melanogenic function. *Journal of Biological Chemistry* **269**, 29198-205.

Kobayashi T., Urabe K., Winder A., Jimenez-Cervantes C., Imokawa G., Brewington T., Solano F., Garcia-Borron J. & Hearing V. (1994b) Tyrosinase related protein 1 (TRP1) functions as a DHICA oxidase in melanin biosynthesis. *The EMBO journal* **13**, 5818.

Kobayashi T., Vieira W.D., Potterf B., Sakai C., Imokawa G. & Hearing V.J. (1995) Modulation of melanogenic protein expression during the switch from eu- to pheomelanogenesis. *Journal of Cell Science* **108**, 2301-9.

Kondo T. & Hearing V.J. (2011) Update on the regulation of mammalian melanocyte function and skin pigmentation. *Expert Review of Dermatology* **6**, 97-108.

Korner A. & Pawelek J. (1982) Mammalian tyrosinase catalyzes three reactions in the biosynthesis of melanin. *Science* **217**, 1163-5.

Korytowski W., Sarna T. & Zareba M. (1995) Antioxidant action of neuromelanin: the mechanism of inhibitory effect on lipid peroxidation. *Archives of Biochemistry and Biophysics* **319**, 142-8.

Kumasaka M., Sato S., Yajima I. & Yamamoto H. (2003) Isolation and Developmental Expression of Tyrosinase Family Genes in Xenopus laevis. *Pigment Cell Research* **16**, 455-62.

Kuroda T.S., Ariga H. & Fukuda M. (2003) The Actin-Binding Domain of Slac2-a/Melanophilin Is Required for Melanosome Distribution in Melanocytes. *Molecular Cell Biology* **23**, 5245-55.

Kurtz S., Phillippy A., Delcher A.L., Smoot M., Shumway M., Antonescu C. & Salzberg S.L. (2004) Versatile and open software for comparing large genomes. *Genome Biology* **5**, 1.

Kushimoto T., Basrur V., Valencia J., Matsunaga J., Vieira W.D., Ferrans V.J., Muller J., Appella E. & Hearing V.J. (2001) A model for melanosome biogenesis based on the purification and analysis of early melanosomes. *Proceedings of the National Academy of Sciences* **98**, 10698-703.

Kwon B.S., Halaban R., Ponnazhagan S., Kim K., Chintamaneni C., Bennett D. & Pickard R.T. (1995) Mouse silver mutation is caused by a single base insertion in the putative cytoplasmic domain of Pmel 17. *Nucleic Acid Research* **23**, 154-8.

Lamoreux L.M., Wakamatsu K. & Ito S. (2001) Interaction of Major Coat Color Gene Functions in Mice as Studied by Chemical Analysis of Eumelanin and Pheomelanin. *Pigment Cell Research* **14**, 23-31.

Le Pape E., Passeron T., Giubellino A., Valencia J.C., Wolber R. & Hearing V.J. (2009) Microarray analysis sheds light on the dedifferentiating role of agouti signal protein in murine melanocytes via the Mc1r. *Proceedings of the National Academy of Sciences* **106**, 1802-7.

Le Pape E., Wakamatsu K., Ito S., Wolber R. & Hearing V.J. (2008) Regulation of eumelanin/pheomelanin synthesis and visible pigmentation in melanocytes by ligands of the melanocortin 1 receptor. *Pigment Cell & Melanoma Research* **21**, 477-86.

Levy C., Khaled M. & Fisher D.E. (2006) MITF: master regulator of melanocyte development and melanoma oncogene. *Trends in Molecular Medicine* **12**, 406-14.

Li H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-93.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. & Subgroup G.P.D.P. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9.

Linnaeus C. (1758) Systema naturae, vol. 1. *Systema naturae, Vol. 1*.

Lister J.A., Robertson C.P., Lepage T., Johnson S.L. & Raible D.W. (1999) nacre encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. *Development* **126**, 3757-67.

Liu J.Z., Mcrae A.F., Nyholt D.R., Medland S.E., Wray N.R., Brown K.M., Hayward N.K., Montgomery G.W., Visscher P.M. & Martin N.G. (2010) A versatile gene-based test for genome-wide association studies. *The American Journal of Human Genetics* **87**, 139-45.

Locke M., Penedo M., Bricker S., Millon L. & Murray J. (2002) Linkage of the grey coat colour locus to microsatellites on horse chromosome 25. *Animal Genetics* **33**, 329-37.

Loman N.J., Misra R.V., Dallman T.J., Constantinidou C., Gharbia S.E., Wain J. & Pallen M.J. (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nature biotechnology* **30**, 434-9.

Australian Alpaca Fleece Limited (2012) Colour Breakdown by Sex and Year Letter. Mitcham, Vic 3132.

Ludwig A., Rehberg S. & Wegner M. (2004) Melanocyte-specific expression of dopachrome tautomerase is dependent on synergistic gene activation by the Sox10 and Mitf transcription factors. *FEBS Letters* **556**, 236-44.

Lupton C.J., McColl A. & Stobart R.H. (2006) Fiber characteristics of the Huacaya Alpaca. *Small Ruminant Research* **64**, 211-24.

Manceau M., Domingues V.S., Mallarino R. & Hoekstra H.E. (2011) The Developmental Role of Agouti in Color Pattern Evolution. *Science* **331**, 1062-5.

Manga P., Sato K., Ye L., Beermann F., Lamoreux M.L. & Orlow S.J. (2000) Mutational Analysis of the Modulation of Tyrosinase by Tyrosinase-Related Proteins 1 and 2 In Vitro. *Pigment Cell Research* **13**, 364-74.

Mariat D., Taourit S. & Guerin G. (2003) A mutation in the MATP gene causes the cream coat colour in the horse. *Genetics Selection Evolution* **35**, 119-33.

Marklund L., Moller M., Sandberg K. & Andersson L. (1996) A missense mutation in the gene for melanocyte-stimulating hormone receptor (MCIR) is associated with the chestnut coat color in horses. *Mammalian Genome* **7**, 895-9.

Marks M.S. & Seabra M.C. (2001) The melanosome: membrane dynamics in black and white. *Nature Reviews Molecular Cell Biology* **2**, 738-48.

Martinez-Arias R., Comas D., Andres A., Abello M.-T., Domingo-Roura X. & Bertranpetit J. (2000) The Tyrosinase Gene in Gorillas and the Albinism of 'Snowflake'. *Pigment Cell Research* **13**, 467-70.

Mártinez-García M. & Montoliu L. (2013) Albinism in Europe. *The Journal of Dermatology* **40**, 319-24.

Mason H., Fowlks W. & Peterson E. (1955) Oxygen Transfer and Electron Transport by the Phenolase Complex1. *Journal of the American Chemical Society* **77**, 2914-5.

Mau C., Poncet P., Bucher B., Stranzinger G. & Rieder S. (2004) Genetic mapping of dominant white (W), a homozygous lethal condition in the horse (Equus caballus). *Journal of Animal Breeding and Genetics* **121**, 374-83.

Mayer T.C. (1973) The migratory pathway of neural crest cells into the skin of mouse embryos. *Developmental Biology* **34**, 39-46.

Mayer T.C. & Green M.C. (1968) An experimental analysis of the pigment defect caused by mutations at the W and Sl loci in mice. *Developmental Biology* **18**, 62-75.

McConlogue L., Brow M.A.D. & Innis M.A. (1988) Structure-independent DNA amplification by PCR using 7-deaza-2'-deoxyguanosine. *Nucleic Acids Research* **16**, 9869-.

McGettigan P.A. (2013) Transcriptomics in the RNA-seq era. *Current Opinion in Chemical Biology* **17**, 4-11.

McGregor B.A. (2006) Production, attributes and relative value of alpaca fleeces in southern Australia and implications for industry development. *Small Ruminant Research* **61**, 93-111.

McGregor B.A., Ramos H.E. & Quispe PeÃ±a E.C. (2011) Variation of fibre characteristics among sampling sites for Huacaya alpaca fleeces from the High Andes. *Small Ruminant Research*.

Mengoni Gonalons G.L. (2008) Camelids in ancient Andean societies: A review of the zooarchaeological evidence. *Quaternary International* **185**, 59-68.

Meunier J. & Duret L. (2004) Recombination drives the evolution of GC-content in the human genome. *Molecular Biology and Evolution* **21**, 984-90.

Miller M.W., Duhl D., Vrieling H., Cordes S.P., Ollmann M.M., Winkes B.M. & Barsh G.S. (1993) Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the lethal yellow mutation. *Genes & Development* **7**, 454-67.

Miller S.A., Dykes D.D. & Polesky H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* **16**, 1215.

Miyamura Y., Coelho S.G., Wolber R., Miller S.A., Wakamatsu K., Zmudzka B.Z., Ito S., Smuda C., Passeron T., Choi W., Batzer J., Yamaguchi Y., Beer J.Z. & Hearing V.J. (2007) Regulation of human skin pigmentation and responses to ultraviolet radiation. *Pigment Cell Research* **20**, 2-13.

Molina G. (1782) Saggio sulla storia naturale del Chile. Bologna. *8º, premiere edition* **367**.

Moller M.J., Chaudhary R., Hellmen E., Höyheim B., Chowdhary B. & Andersson L. (1996) Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor. *Mammalian Genome* **7**, 822-30.

Morlan J.D., Qu K. & Sinicropi D.V. (2012) Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS One* **7**, e42882.

Morozova O. & Marra M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**, 255-64.

Müller O.F. (1776) *Zoologiae Danicae prodromus: seu Animalium Daniae et Norvegiae indigenarum characteres, nomina, et synonyma imprimis popularium*. typis Hallageriis.

Munyard K. (2008) The Search for Molecular Markers in Alpacas: Genes that Influence Colour. In: *World Alpaca Conference*. Australian Alpaca Association, Sydney.

Munyard K. & Cransberg R. (2016) The Expression Level of the Alpaca Agouti Gene has a Marked Impact on Fibre Colour Intensity. In: *International Plant & Animal Genome XXIV*, San Diego, CA.

Munyard K.A. (2011) *Inheritance of White Colour in Alpacas, Identifying the genes involved*. Rural Industries Research and Development Corporation, Barton, ACT.

Nadeau N.J., Minvielle F., Ito S.i., Inoue-Murayama M., Gourichon D., Follett S.A., Burke T. & Mundy N.I. (2008) Characterization of Japanese quail yellow as a genomic deletion upstream of the avian homolog of the mammalian ASIP (agouti) gene. *Genetics* **178**, 777-86.

Najamabadi M. (1934) Tarikh-e-Tib-e-Iran. *Vol I. Shamsi: Tehran*.

Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davisson M.T., King R.A. & Brilliant M.H. (2001) Mutations in the Human Orthologue of the Mouse underwhite Gene (uw) Underlie a New Form of Oculocutaneous Albinism, OCA4. *The American Journal of Human Genetics* **69**, 981-8.

Newton J.M., Wilkie A.L., He L., Jordan S.A., Metallinos D.L., Holmes N.G., Jackson I.J. & Barsh G.S. (2000) Melanocortin 1 receptor variation in the domestic dog. *Mammalian Genome* **11**, 24-30.

Nicolaus B.J. (2005) A critical review of the function of neuromelanin and an attempt to provide a unified theory. *Medical Hypotheses* **65**, 791-6.

Norris B.J. & Whan V.A. (2008) A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Research* **18**, 1282-93.

O'Sullivan T.N., Wu X.S., Rachel R.A., Huang J.-D., Swing D.A., Matesic L.E., Hammer J.A., Copeland N.G. & Jenkins N.A. (2004) dsu functions in a MYO5A-independent pathway to suppress the coat color of dilute mice. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16831-6.

O'Brien S.J., Pontius J.U., Johnson W.E. & Perelman P.L. (2008) The Alpaca Enters The Genomic Era In: *1st International Workshop on Camelid Genetics*, Scottsdale, Arizona.

Oyehaug L., Plahte E., Vage D.I. & Omholt S.W. (2002) The Regulatory Basis of Melanogenic Switching. *Journal of Theoretical Biology* **215**, 449-68.

Ozeki H., Ito S., Wakamatsu K. & Hirobe T. (1995) Chemical Characterization of Hair Melanins in Various Coat-Color Mutants of Mice. *Journal of Investigative Dermatology* **105**, 361-6.

Ozeki H., Ito S., Wakamatsu K. & Ishiguro I. (1997a) Chemical characterization of pheomelanogenesis starting from dihydroxyphenylalanine or tyrosine and cysteine.: Effects of tyrosinase and cysteine concentrations and reaction time. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1336**, 539-48.

Ozeki H., Ito S., Wakamatsu K. & Thody A.J. (1996) Spectrophotometric Characterization of Eumelanin and Pheomelanin in Hair. *Pigment Cell Research* **9**, 265-70.

Ozeki H., Wakamatsu K., Ito S. & Ishiguro I. (1997b) Chemical Characterization of Eumelanins with Special Emphasis on 5,6-Dihydroxyindole-2-carboxylic Acid Content and Molecular Size. *Analytical Biochemistry* **248**, 149-57.

Paul E. (2002) *The Alpaca Colour Key*. Paul Publications, Perth, Australia.

Paul E. (2006) Colour Review. In: *Australian Alpaca Association National Conference*, Adelaide.

Pedersen B.S. (2011) solidtrimmer.py.

Philipp U., Hamann H., Mecklenburg L., Nishino S., Mignot E., Günzel-Apel A.-R., Schmutz S.M. & Leeb T. (2005) Polymorphisms within the canine MLPH gene are associated with dilute coat color in dogs. *BMC Genetics* **6**, 34.

Pinard R., de Winter A., Sarkis G.J., Gerstein M.B., Tartaro K.R., Plant R.N., Egholm M., Rothberg J.M. & Leamon J.H. (2006) Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics* **7**, 216.

Pomp D. & Medrano J. (1991) Organic solvents as facilitators of polymerase chain reaction. *Biotechniques* **10**, 58-9.

Potterf S.B., Furumura M., Dunn K.J., Arnheiter H. & Pavan W.J. (2000) Transcription factor hierarchy in Waardenburg syndrome: regulation of MITF expression by SOX10 and PAX3. *Human Genetics* **107**, 1-6.

Potterf S.B., Mollaaghababa R., Hou L., Southard-Smith E.M., Hornyak T.J., Arnheiter H. & Pavan W.J. (2001) Analysis of SOX10 Function in Neural Crest-Derived Melanocyte Development: SOX10-Dependent Transcriptional Control of Dopachrome Tautomerase. *Developmental Biology* **237**, 245-57.

Powell A.J., Moss M.J., Tree L.T., Roeder B.L., Carleton C.L., Campbell E. & Kooyman D.L. (2008) Characterization of the effect of Melanocortin 1 Receptor, a member of the hair color genetic locus, in alpaca (Lama pacos) fleece color differentiation. *Small Ruminant Research* **79**, 183-7.

Prota G. (1993) Regulatory mechanisms of melanogenesis: beyond the tyrosinase concept. *Journal of Investigative Dermatology* **100**.

Rajora N., Ceriani G., Catania A., Star R.A., Murphy M.T. & Lipton J.M. (1996) alpha-MSH production, receptors, and influence on neopterin in a human monocyte/macrophage cell line. *Journal of Leukocyte Biology* **59**, 248-53.

Rawles M.E. (1947) Origin of Pigment Cells from the Neural Crest in the Mouse Embryo. *Physiological Zoology* **20**, 248-66.

Reissmann M., Bierwolf J. & Brockmann G.A. (2007) Two SNPs in the SILV gene are associated with silver coat colour in ponies. *Animal Genetics* **38**, 1-6.

Rhoads A. & Au K.F. (2015) PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics* **13**, 278-89.

Rieder S., Taourit S., Mariat D., Langlois B. & Guérin G. (2001) Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (Equus caballus). *Mammalian Genome* **12**, 450-5.

Rinchik E.M., Bultman S.J., Horsthemke B., Lee S.-T., Strunk K.M., Spritz R.A., Avidano K.M., Jong M.T. & Nicholls R.D. (1993) A gene for the mouse pink-eyed dilution locus and for human type II oculocutaneous albinism. *Nature* **361**, 72-6.

Rosemblat S., Durham-Pierre D., Gardner J.M., Nakatsu Y., Brilliant M.H. & Orlow S.J. (1994) Identification of a melanosomal membrane protein encoded by the pink-eyed dilution (type II oculocutaneous albinism) gene. *Proceedings of the National Academy of Sciences* **91**, 12071-5.

Saito H., Yasumoto K.-I., Takeda K., Takahashi K., Yamamoto H. & Shibahara S. (2003) Microphthalmia-Associated Transcription Factor in the Wnt Signaling Pathway. *Pigment Cell Research* **16**, 261-5.

Sato-Jin K., Nishimura E.K., Akasaka E., Huber W., Nakano H., Miller A., Du J., Wu M., Hanada K., Sawamura D., Fisher D.E. & Imokawa G. (2008) Epistatic connections between microphthalmia-associated transcription factor and endothelin signaling in Waardenburg syndrome and other pigmentary disorders. *The FASEB Journal* **22**, 1155-68.

Sato S., Reiko T., You K., Hidetoshi S., Takaharu N., Kazuho I., Takashi G., Ichiro Y. & Hiroaki Y. (1999) Structure and developmental expression of the ascidian TRP gene: Insights into the evolution of pigment cell-specific gene expression. *Developmental Dynamics* **215**, 225-37.

Schmidt D., Wilson M.D., Spyrou C., Brown G.D., Hadfield J. & Odom D.T. (2009) ChIP-seq: using high-throughput sequencing to discover protein–DNA interactions. *Methods* **48**, 240-8.

Schmutz S.M. & Berryere T.G. (2007) Genes affecting coat colour and pattern in domestic dogs: a review. *Animal Genetics* **38**, 539-49.

Schmutz S.M., Berryere T.G., Ciobanu D.C., Mileham A.J., Schmidtz B.H. & Fredholm M. (2004) A form of albinism in cattle is caused by a tyrosinase frameshift mutation. *Mammalian Genome* **15**, 62-7.

Schmutz S.M., Berryere T.G. & Goldfinch A.D. (2002) TYRP1 and MC1R genotypes and their effects on coat color in dogs. *Mammalian Genome* **13**, 380-7.

Secundus G.P. & Jones W.H.S. (1980) *Natural History*. Harvard University Press, Cambridge, MA.

Shendure J. & Ji H. (2008) Next-generation DNA sequencing. *Nature Biotechnology* **26**, 1135-45.

Shokralla S., Spall J.L., Gibson J.F. & Hajibabaei M. (2012) Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* **21**, 1794-805.

Simon G. (1841) Zur Entwicklung des Haares. *joh. müflers arch.*

Simpson Q. (1914) Coat-Pattern in Mammals A Medium of Real Value to the Breeder, Since it Enables Him by Analysis to Detect in Many Cases the Genetic Composition of His Animals—Mendelism in the Hands of the Fancier. *Journal of Heredity* **5**, 329-39.

Sitaram A. & Marks M.S. (2012) Mechanisms of Protein Delivery to Melanosomes in Pigment Cells. *Physiology* **27**, 85-99.

Slominski A., Tobin D.J., Shibahara S. & Wortsman J. (2004) Melanin Pigmentation in Mammalian Skin and Its Hormonal Regulation. *Physiological Reviews* **84**, 1155-228.

Souvorov A., Kapustin Y., Kiryutin B., Chetvernin V., Tatusova T. & Lipman T. (2010). GNOMON-NCBI eukaryotic gene prediction tool. *National Center for Biotechnology Information*, 1-24

Sponenberg D.P., Ito S., Eng L.A. & Schwink K. (1988) Pigment types of various color genotypes of horses. *Pigment Cell Research* **1**, 410-3.

Steiner C.C., Weber J.N. & Hoekstra H.E. (2007) Adaptive Variation in Beach Mice Produced by Two Interacting Pigmentation Genes. *PLoS Biology* **5**, e219.

Sturm R.A. & Duffy D.L. (2012) Human pigmentation genes under environmental selection. *Genome Biology* **13**, 1.

Sulaimon S.S. & Kitchell B.E. (2003) The biology of melanocytes. *Veterinary Dermatology* **14**, 57-65.

Szklarczyk D., Franceschini A., Wyder S., Forslund K., Heller D., Huerta-Cepas J., Simonovic M., Roth A., Santos A. & Tsafou K.P. (2014) STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, gku1003.

Thiruvenkadan A.K., Kandasamy N. & Panneerselvam S. (2008) Coat colour inheritance in horses. *Livestock Science* **117**, 109-29.

Tian X., Meng X., Wang L., Song Y., Zhang D., Ji Y., Li X. & Dong C. (2015) Molecular cloning, mRNA expression and tissue distribution analysis of Slc7a11 gene in alpaca (Lama paco) skins associated with different coat colors. *Gene* **555**, 88-94.

Tobin D.J. (2011) The cell biology of human hair follicle pigmentation. *Pigment Cell and Melanoma Research* **24**, 75-88.

Tolleson W.H. (2005) Human melanocyte biology, toxicology, and pathology. *Journal of Environmental Science and Health Part C* **23**, 105-61.

Tomita Y., Takeda A., Okinaga S., Tagami H. & Shibahara S. (1989) Human oculocutaneous albinism caused by single base insertion in the tyrosinase gene. *Biochemical and Biophysical Research Communications* **164**, 990-6.

Trapnell C., Pachter L. & Salzberg S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11.

Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., van Baren M.J., Salzberg S.L., Wold B.J. & Pachter L. (2010) Transcript assembly and quantification

by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**, 511-5.

Tsukamoto K., Palumbo A., d'Ischia M., Hearing V. & Prota G. (1992) 5, 6-Dihydroxyindole-2-carboxylic acid is incorporated in mammalian melanin. *Biochemical Journal* **286**, 491-5.

Tully G. (2007) Genotype versus phenotype: Human pigmentation. *Forensic Science International: Genetics* **1**, 105-10.

Vachtenheim J. & Borovanský J. (2010) "Transcription physiology" of pigment formation in melanocytes: central role of MITF. *Experimental Dermatology* **19**, 617-27.

Valbonesi A., Apaza N., La Manna V., Gonzales M.L., Huanca T. & Renieri C. (2011) Inheritance of white, black and brown coat colours in alpaca (Vicuna pacos L.). *Small Ruminant Research* **99**, 16-9.

Visscher P.M., Brown M.A., McCarthy M.I. & Yang J. (2012) Five years of GWAS discovery. *The American Journal of Human Genetics* **90**, 7-24.

Vrieling H., Duhl D., Millar S.E., Miller K.A. & Barsh G.S. (1994) Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. *Proceedings of the National Academy of Sciences* **91**, 5667-71.

Wakamatsu K., Fujikawa K., Zucca F.A., Zecca L. & Ito S. (2003) The structure of neuromelanin as studied by chemical degradative methods. *Journal of Neurochemistry* **86**, 1015-23.

Wakamatsu K. & Ito S. (2002) Advanced Chemical Methods in Melanin Determination. *Pigment Cell Research* **15**, 174-83.

Wakamatsu K., Ohtara K. & Ito S. (2009) Chemical analysis of late stages of pheomelanogenesis: conversion of dihydrobenzothiazine to a benzothiazole structure. *Pigment Cell & Melanoma Research* **22**, 474-86.

Wang N. & Hebert D.N. (2006) Tyrosinase maturation through the mammalian secretory pathway: bringing color to life. *Pigment Cell Research* **19**, 3-18.

Wang Z., Gerstein M. & Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57-63.

Wehrle-Haller B. (2003) The Role of Kit-Ligand in Melanocyte Development and Epidermal Homeostasis. *Pigment Cell Research* **16**, 287-96.

Wehrle-Haller B. & Weston J.A. (1999) Altered cell-surface targeting of stem cell factor causes loss of melanocyte precursors in Steel(17H) mutant mice. *Developmental Biology* **210**, 71-86.

Weikard R., Hadlich F. & Kuehn C. (2013) Identification of novel transcripts and noncoding RNAs in bovine skin by deep next generation sequencing. *BMC Genomics* **14**, 1.

Westerhof W. (2006) The discovery of the human melanocyte. *Pigment Cell Research* **19**, 183-93.

Wheeler J.C. (1995) Evolution and present situation of the South American Camelidae. *Biological Journal of the Linnean Society* **54**, 271-95.

Wheeler J.C., Russel A.J.F. & Redden H. (1995) Llamas and Alpacas: Pre-conquest breeds and post-conquest hybrids. *Journal of Archaeological Science* **22**, 833-40.

Widlund H.R. & Fisher D.E. (2003) Microphthalamia-associated transcription factor: a critical regulator of pigment cell development and survival. *Oncogene* **22**, 3035-41.

Wilczek A., Kondoh H. & Mishima Y. (1996) Composition of Mammalian Eumelanins: Analyses of DHICA-Derived Units in Pigments From Hair and Melanoma Cells. *Pigment Cell Research* **9**, 63-7.

Wilson S.M., Yip R., Swing D.A., O'Sullivan T.N., Zhang Y., Novak E.K., Swank R.T., Russell L.B., Copeland N.G. & Jenkins N.A. (2000) A mutation in Rab27a causes the vesicle transport defects observed in ashen mice. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 7933-8.

Wright S. (1917) Color Inheritance in Mammals. *Journal of Heredity* **8**, 224-35.

Wu X., Rao K., Bowers M.B., Copeland N.G., Jenkins N.A. & Hammer J.A. (2001) Rab27a enables myosin Va-dependent melanosome capture by recruiting the myosin to the organelle. *Journal of Cell Science* **114**, 1091-100.

Yajima I., Sato S., Kimura T., Yasumoto K.-i., Shibahara S., Goding C.R. & Yamamoto H. (1999) An L1 Element Intronic Insertion in the Black-Eyed White (Mitfmi-bw) Gene: The Loss of a Single Mitf Isoform Responsible for the Pigmentary Defect and Inner Ear Deafness. *Human Molecular Genetics* **8**, 1431-41.

Yamaguchi Y., Brenner M. & Hearing V.J. (2007) The Regulation of Skin Pigmentation. *Journal of Biological Chemistry* **282**, 27557-61.

Yamaguchi Y. & Hearing V.J. (2009) Physiological factors that regulate skin pigmentation. *Biofactors* **35**, 193-9.

Yang G., Li Y., Nishimura E.K., Xin H., Zhou A., Guo Y., Dong L., Denning M.F., Nickoloff B.J. & Cui R. (2008) Inhibition of PAX3 by TGF-[beta] Modulates Melanocyte Viability. *Molecular Cell* **32**, 554-63.

Yang S., Fan R., Shi Z., Ji K., Zhang J., Wang H., Herrid M., Zhang Q., Yao J. & Smith G. (2015) Identification of a novel microRNA important for melanogenesis in alpaca (). *Journal of Animal Science* **93**, 1622-31.

Zaręba M., Bober A., Korytowski W., Zecca L. & Sarna T. (1995) The effect of a synthetic neuromelanin on yield of free hydroxyl radicals generated in model systems. *Biochimica et biophysica acta* **1271**, 343-8.

Zdarsky E., Favor J. & Jackson I.J. (1990) The Molecular Basis of brown, an Old Mouse Mutation, and of an Induced Revertant to Wild Type. *Genetics* **126**, 443-9.

Zhang J., Thomas D., Suri P. & Yang C. (2011) XSQ: A New Binary Output Format of 5500/5500XL Systems. *Journal of Biomolecular Techniques : JBT* **22**, S28-S9.

Zhao S., Fung-Leung W.-P., Bittner A., Ngo K. & Liu X. (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* **9**, e78644.

Zhu Z., He J., Jia X., Jiang J., Bai R., Yu X., Lv L., Fan R., He X. & Geng J. (2010) MicroRNA-25 functions in regulation of pigmentation by targeting the transcription factor MITF in alpaca (Lama pacos) skin melanocytes. *Domestic Animal Endocrinology* **38**, 200-9.

Appendices

Please access the appendices via the following link:

https://cloudstor.aarnet.edu.au/plus/index.php/s/dWWXqjl2MaeQ8yL