

1 **Title** Polymorphism of sheep MHC Class IIb gene *TAPASIN*

2

3

4

5 **Authors** N. Siva Subramaniam, E. F. Morgan, C. Y. Lee, J.D. Wetherall and D.M.
6 Groth

7

8 Western Australian Biomedical Research Institute (WABRI) & Centre for Health
9 Innovation Research Institute, School of Biomedical Sciences, Curtin University of
10 Technology, Perth, Western Australia, 6845

11

12

13 **Correspondence** Associate Professor David Groth, School of Biomedical Sciences,
14 Curtin University, GPO Box U1987, Perth, Western Australia,
15 6845.

16

Fax: 61892662342

17

Email: D.Groth@curtin.edu.au

18

19

20

21 **Abstract**

22

23 The Major Histocompatibility Complex (MHC) is one of the most gene dense regions
24 in the genome and studies in several species have shown significant associations
25 between the MHC and disease. The endoplasmic reticular glycoprotein, tapasin, is
26 involved in the MHC class I antigen presentation pathway. Sheep *TAPASIN* is located
27 in the class IIb region of the MHC. Sheep *TAPASIN* was subcloned from BAC and
28 cosmid genomic clones and DNA sequenced. *TAPASIN* is 9549 bp in length and
29 encodes a protein of 447 amino acids. The structure of sheep *TAPASIN* was similar to
30 other mammals and consisted of eight exons with a distinctively larger intron between
31 exon three and four. Sheep *TAPASIN* gene had high sequence identity with other
32 mammalian *TAPASINs*. The *TAPASIN* gene sequence is conserved across many
33 mammalian species and is possibly maintained through purifying selection with the
34 average ratio of ds/dn of 3.9. Twenty-six SNPs in sheep *TAPASIN* were identified.

35

36 **Keywords** *TAPASIN*, TAP-binding protein, MHC, sheep, SNP, polymorphism
37 Tapasin is an MHC class IIb region encoded (*TAPASIN*) protein involved in the
38 classical class I antigen presentation pathway. Tapasin, also known as transporter
39 associated with antigen processing binding protein (TAPBP), is an endoplasmic
40 reticular glycoprotein (Garbi et al. 2003). In the class I presentation pathway, tapasin
41 has several important functions such as the recruitment of transporters associated with
42 antigen processing protein (TAP) in the endoplasmic reticulum (ER) (Lehner et al.
43 1998), stabilisation of TAP and class I molecule interactions required for optimal
44 peptide loading (Ortmann et al. 1997), assembly of the class I heavy chain (Garbi et
45 al. 2000), peptide selection in the peptide loading complex (PLC) (Garbi et al. 2000;

46 Howarth et al. 2004) and retention of the class I molecules in the endoplasmic
47 reticulum (Grande III *et al.* 1995; Grande III *et al.* 2000; Grande III & Van Kaer
48 2001). Mutations within the *TAPASIN* gene can significantly disrupt the functional
49 role of tapasin in the MHC class I presentation pathway (Copeman et al. 1998). The
50 “loss of function” phenotype results in a decrease of antigen presentation at the cell
51 surface (Copeman et al. 1998). The *TAPASIN* gene in sheep is located within the class
52 IIb region of the sheep MHC on chromosome 20 (Mahdy et al. 1989). Unlike humans
53 and most other mammals, the class II region in sheep is split into IIa and IIb regions
54 separated by a non-MHC region (Liu et al. 2006).

55

56 Several studies have shown direct or indirect associations between the sheep MHC
57 and Quantitative Trait Loci (QTL) for resistance to disease. In Scottish blackface
58 sheep naturally infested with *Ostertagia circumcincta*, certain MHC class II antigens
59 were associated with 98% lower egg count (Schwaiger et al. 1995). A recent study in
60 Scottish blackface sheep aimed at identifying QTLs for a variety of parasite resistance
61 indicators showed that there was a significant chromosome wide QTL located within
62 the class IIb region (Davies et al. 2006). Significant associations between MHC and
63 intestinal nematodes have been reported in several other studies (Outteridge et al.
64 1996; Paterson et al. 1998; van Haeringen et al. 1999). Recently, microarray analysis
65 of the response of Perendale sheep to nematodes showed that the more resistant lambs
66 had higher expression of the MHC class II genes (Diez-Tascon et al. 2005). In
67 Rhonschaf sheep, significant associations between faecal egg counts (FEC) and the
68 markers *OarCp73*, *DYMS1* and *BM1815* was observed. The *DYA* gene located within
69 the class IIb subregion of the MHC is closely linked to the microsatellite *DYMS* and

70 is a possible candidate gene for conferring resistance to *Haemonchus contortus* in
71 these sheep (Charon 2004).

72 This report describes the structure, DNA sequence and single nucleotide
73 polymorphisms (SNP) of *TAPASIN*, which is located within the sheep MHC Class IIb
74 gene region on chromosome 20. An analysis of the ratio of synonymous to non-
75 synonymous substitutions in the coding region of this gene across several mammalian
76 species was also performed. This study is part of an ongoing development of a
77 haplotypic analysis of the sheep MHC class II region. Information from this study will
78 assist in understanding the important role genes within the MHC play in immunity to
79 disease.

80 Genomic DNA was isolated from the leukocytes from individual merino sheep using
81 a Qiagen tissue DNA isolation kit according to manufacturer's instructions. DNA
82 was suspended in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) and stored at -20°C
83 until required. *TAPASIN* degenerative PCR primers
84 (5'CTGYCTTGYRTCCCACTTCT3' and 5'CCAGGGTGACCTCAGCRCTG 3')
85 were identified from the alignment of human and mouse *TAPASIN* gene sequences.
86 These primers amplify a small 259bp DNA fragment of the sheep *TAPASIN* gene
87 (Qin 2009) that could be sequenced and from which overgo primers (Gustafson et al.
88 2003) were designed (5'TGCAGAGAGGCTTACAGAGCCATC3' and
89 5'CCTGAGACATCACTCAGATGGCTC3'). The PCR reaction mix (50 ul)
90 comprised: 100 ng sheep genomic DNA, 1 X PCR buffer (Invitrogen), 1.5 mM
91 MgCl₂, 200 uM dNTP, 5 pmol of each primer (Geneworks), 0.2 mg/ml BSA (Roche)
92 and 2.5U Platinum Taq polymerase (Invitrogen). The Overgo primers were
93 radioactively labelled using overgo technology (Gustafson et al., 2003) and hybridised
94 to BAC (CHORI 243) genomic DNA library filters in a buffered solution containing

95 20 × SSPE, 1%BSA, 7%SDS and 0.5M EDTA. The filters were incubated at 53°C
96 overnight, and washed in 1xSSC buffer containing 0.1% SDS. The filters were sealed
97 in plastic and exposed to Kodak x-ray film with an intensifying screens at -80°C for 2
98 days prior to development of the film. DNA from the CHORI clone 461K3 clone was
99 isolated using Qiagen large construct kit according to the manufacturer's protocol and
100 digested with Pst I and cloned into pGEM5f. Random clones were selected and
101 sequenced using M13 primers by Macrogen (Korea). Clones containing *TAPASIN*
102 sequences were subjected to further sequencing.

103 Multiple pass DNA sequencing resulted in approximately 11 Kbp of quality sequence,
104 with *TAPASIN* identified as being 9549 bp in length and encoding a predicted protein
105 of 447 amino acids. The overall structure of the sheep *TAPASIN* gene (Genbank
106 accession EU814901) was similar to that reported for several other mammals, cattle
107 (NW_001494145), chimpanzee (NW_001236523), dog (NW_876254), human
108 (NW_923073) and mouse (NW_001030615). A consensus genetic structure for sheep
109 *TAPASIN* was determined from analysis of the predictions from GENSCAN (Burge &
110 Karlin 1997) Burge and Karlin, 1997), TWINSCAN (Korf et al. 2001) and
111 FGENESH (<http://www.softberry.com>). Sheep *TAPASIN* was found to comprise eight
112 exons, with a distinctively larger intron of 5526 bp found between exons three and
113 four.

114 Clustal W (Thompson et al. 1994) multiple sequence alignment was performed with
115 *TAPASIN* from sheep, cattle, human, rat, mouse, horse, chimpanzee, macaca, dog and
116 zebrafish. This analysis showed that there was significant amino acid sequence
117 conservation (Figure 1). The various features of tapasin including signal peptide, C
118 and V immunoglobulin domains, conserved cysteine residues, motifs 3 and 4,
119 transmembrane region and cytoplasmic tail could all be identified and are also shown

120 in Figure 1. These features were identical to those shown previously for rat, Atlantic
121 salmon and several other mammals (Ortmann *et al.* 1997; Deverson *et al.* 2001;
122 Jorgensen *et al.* 2007). The average DNA sequence identity between the coding
123 sequence of sheep *TAPASIN* gene and other organisms was 85.3% whereas the amino
124 acid sequence identity was slightly lower at 83.6%. BLAST sequence alignment
125 showed higher DNA sequence identity with cattle *TAPASIN* (95%) when compared to
126 a similar analysis with human *TAPASIN* (84%). A bootstrapped Neighbour-Joining
127 phylogenetic tree was constructed using the Clustal W multiple sequence alignment.
128 100 percent of the trees obtained placed the sheep and cattle tapasin in a separate
129 clade from the primate tapasins (1000 bootstraps), indicating, as expected, that sheep
130 tapasin is more closely related to cattle tapasin than to the primate tapasins (Sup.
131 Figure 1). Relative to the cattle amino acid sequence, there are amino acid differences
132 in the IgV domain (positions 170 and 171). However the observed change at position
133 170 appears to predate the generation of the ungulate lineage as it is also present in
134 primates, horses and zebrafish. The variation position 171 however appears to be
135 unique to sheep, at least in this small cohort of species. In the TPN motif 3 domain,
136 there is a deleted amino acid in cattle relative to the other species. This position in
137 sheep is the identical to horse *TAPASIN* implying a three base pair deletion has
138 occurred in cattle. Unique amino acid changes occur at position 4 in the leader
139 sequence and position 323 in the Ig C domain.

140 Synonymous and non-synonymous nucleotide substitutions were analysed using the
141 SNAP software package (www.hiv.lanl.gov and
142 <http://hcv.lanl.gov/content/sequence/SNAP/SNAP.html>) (Nei & Gojobori 1986;
143 Korber 2000). This analysis showed that the *TAPASIN* coding sequence is conserved
144 across all the mammalian species studied. The average ratio of Jukes-Cantor corrected

145 ds/dn was 3.91 (ranging from Jukes-Cantor corrected ds/dn of 2.48 for chimpanzee
146 versus human to a Jukes-Cantor corrected ds/dn of 4.88 for sheep versus human). This
147 high conservation is most likely maintained through purifying selection. Tapasin is an
148 important component of the MHC class I assembly with several domains shown to
149 have critical functions in the assembly process (Rizvi & Raghavan 2009). Therefore,
150 it is not surprising that tapasin displays high amino acid sequence conservation
151 between species. The amino acid sequence comparisons are shown in supplementary
152 Figure 2.

153 Sequencing was performed on six unrelated merino sheep. Pairwise alignment of
154 DNA sequences using Vector NTI software (Invitrogen) resulted in the direct
155 identification of thirteen SNPs. Furthermore, when BLAST alignment of the
156 consensus sheep sequence obtained in this study was performed against the genome
157 sequence data generated by the International Sheep Genome Consortium (ISGC)
158 (<https://isgcdata.agresearch.co.nz/>), thirteen additional SNPs were identified. Each
159 SNP identified from the ISGC data occurred either within different breeds of sheep or
160 multiple sheep within one breed. Seven of the 13 SNP (54%) identified in this project
161 by sequence alignment of the 6 unrelated merino sheep were also independently
162 verified through the BLAST alignment with the ISGC data. The estimated variation
163 within the sheep *TAPASIN* gene is approximately one SNP per 367 bp. The SNP
164 density observed for *TAPASIN* was comparable with the genome wide SNP survey
165 (Kijas et al, 2009). All the SNPs identified in this study were located within non-
166 coding regions of the gene. It was observed however, that the DNA sequence
167 generated in this study when subject to BLAST alignment with the ISGC data, the
168 matches in the ISGC data did not completely span the entire gene, suggesting possible

169 gaps in the ISGC sequence data. Details and position of all the SNPs identified in this
170 study and the ISGC data are shown in Table 1.

171 The SNPs that have been identified in this study will contribute to an essential
172 framework of SNPs in the MHC class IIb region and thus will be an important
173 resource for the future characterisation and dissection of the sheep MHC haplotypes
174 and their possible role in disease.

175

176 **Acknowledgements**

177

178 PhD Studentship support for NSS from WABRI and Curtin University. All animal
179 experiments were performed according to the Australian Code of Practice for the care
180 and use of the animals for scientific purposes. The authors would like to acknowledge
181 the contribution of Dr Brian Dalrymple to the ISGC screening.

182

183 **References**

184

- 185 Burge C & Karlin S (1997) Prediction of complete gene structures in human genomic
186 DNA. *Journal of Molecular Biology* 268: 78-94
- 187 Charon KM (2004) Genes Controlling Resistance to Gastrointestinal Nematodes in
188 Ruminants. *Animal Science Papers and Reports* 22: 135 - 9
- 189 Copeman J, Bangia NJC & Cresswell CP (1998) Elucidation of the genetic basis of
190 the antigen presentation defects in the mutant cell line .220 reveals
191 polymorphism and alternative splicing of the tapasin gene. *European Journal*
192 *of Immunology* 28: 3783-91
- 193 Davies G, Stear MJ, Benothman M, Abuagob O, Kerr A, Mitchell S & Bishop SC
194 (2006) Quantitative trait loci associated with parasitic infection in Scottish
195 blackface sheep. *Heredity* 96: 252
- 196 Deverson EV, Powis SJ, Morrice NA, Herberg JA, Trowsdale J & Butcher GW
197 (2001) Rat tapasin: cDNA cloning and identification as a component of the
198 class I MHC assembly complex. *Genes Immun* 2: 48-51
- 199 Diez-Tascon C, Keane OM, Wilson T, Zadissa A, Hyndman DL, Baird DB, McEwan
200 JC & Crawford AM (2005) Microarray analysis of selection lines from
201 outbred populations to identify genes involved with nematode parasite
202 resistance in sheep. *Physiol. Genomics* 21: 59-69

203 Garbi N, Tan P, Diehl AD, Chambers BJ, Ljunggren HG, Momburg F & Hammerling
204 GJ (2000) Impaired immune responses and altered peptide repertoire in
205 tapasin-deficient mice. *Nature Immunology* 1: 234

206 Garbi N, Tiwari N, Momburg F & Hammerling GJ (2003) A major role for tapasin as
207 a stabilizer of the TAP peptide transporter and consequences for MHC class I
208 expression. *Immunology* 33: 264-73

209 Grandea III AG, Androlewicz MJ, Athwal RS, Geraghty DE & Spies T (1995)
210 Dependence of Peptide Binding by MHC Class I Molecules on Their
211 Interaction with TAP. *Science* 270: 105-8

212 Grandea III AG, Golovina TN, Hamilton SE, Sriram V, Spies T, Brutkiewicz RR,
213 Harty JT, Eisenlohr LC & Van Kaer L (2000) Impaired Assembly yet Normal
214 Trafficking of MHC Class I Molecules in Tapasin Mutant Mice. *Immunity* 13:
215 213-22

216 Grandea III AG & Van Kaer L (2001) Tapasin: an ER chaperone that controls MHC
217 class I assembly with peptide. *Trends in Immunology* 22: 194-9

218 Gustafson AL, Tallmadge RL, Ramlachan N, Miller D, Bird H, Antczak DF,
219 Raudsepp T, Chowdhary BP & Skow LC (2003) An ordered BAC contig map
220 of the equine major histocompatibility complex. *Cytogenetic and Genome*
221 *Research* 102: 189

222 Howarth M, Williams A, Tolstrup AB & Elliott T (2004) Tapasin enhances MHC class
223 I peptide presentation according to peptide half-life. *PNAS* 101: 11737-42

224 Jorgensen SM, Grimholt U & Gjoen T (2007) Cloning and expression analysis of an
225 Atlantic salmon (*Salmo salar* L.) tapasin gene. *Dev Comp Immunol* 31: 708-
226 19

227 Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, et al. (2009) A
228 Genome Wide Survey of SNP Variation Reveals the Genetic Structure of
229 Sheep Breeds. *PLoS ONE* 4(3): e4668. doi:10.1371/journal.pone.0004668

230 Korber B (2000) HIV Signature and Sequence Variation Analysis. *Computational*
231 *Analysis of HIV Molecular Sequences*, Chapter 4, pages 55-72. Allen G.
232 Rodrigo and Gerald H. Learn, eds. Dordrecht, Netherlands: Kluwer Academic
233 Publishers

234 Korf I, Flicek PD, Duan & Brent MR (2001) "Integrating genomic homology into
235 gene structure prediction". *Bioinformatics* 17: 140-8

236 Lehner PJ, Surman MJ & Cresswell P (1998) Soluble Tapasin Restores MHC Class I
237 Expression and Function in the Tapasin-Negative Cell Line .220. *Immunity* 8:
238 221-31

239 Liu H, Liu K, Wang J & Ma RZ (2006) A BAC clone-based physical map of ovine
240 major histocompatibility complex. *Genomics* 88: 88-95

241 Mahdy EA, Makinen A, Chowdhary BP, Andersson L, Gustavsson I & (1989)
242 Chromosomal localization of the ovine major histocompatibility complex
243 (OLA) by in situ hybridization. *Hereditas* 111: 87-90

244 Nei M & Gojobori T (1986) Simple methods for estimating the numbers of
245 synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:
246 418-26

247 Ortmann B, Copeman J, Lehner PJ, Sadasivan B, Herberg JA, Grandea AG & et al
248 (1997) A critical role for tapasin in the assembly and function of multimeric
249 MHC class I-TAP complexes. *Science* 277: 1306-9

250 Outteridge PM, Andersson L, Douch PGC, Green RS, Gwakisa PS, Hohenhaus MA
251 & Mikko S (1996) The PCR typing of MHC-DRB genes in the sheep using

252 primers for an intronic microsatellite: Application to nematode parasite
253 resistance. *Immunol Cell Biol* 74: 330-6
254 Paterson S, Wilson K & Pemberton JM (1998) Major histocompatibility complex
255 variation associated with juvenile survival and parasite resistance in a large
256 unmanaged ungulate population. *Proceedings of the National Academy of
257 Sciences of the United States of America* 95: 3714-9
258 Qin J (2009) Characterisation of the Central Region of the Sheep Major
259 Histocompatibility Complex. Dissertation, Curtin University of Technology,
260 Perth
261 Rizvi SM & Raghavan M (2009) Mechanisms of Function of Tapasin, a Critical
262 Major Histocompatibility Complex Class I Assembly Factor. *Traffic*
263 doi:10.1111/j.1600-0854.2009.01025.x
264 Schwaiger FW, Gostomski D, Stear MJ, Duncan JL, McKellar QA, Epplen JT &
265 Buitkamp J (1995) An ovine Major histocompatibility complex DRB1 allele is
266 associated with low faecal egg counts following natural, predominantly
267 *Ostertagia circumcincta* infection. *International Journal for Parasitology* 25:
268 815-22
269 Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the
270 sensitivity of progressive multiple sequence alignment through sequence
271 weighting, positions-specific gap penalties and weight matrix choice. *Nucleic
272 Acids Research* 22: 4673-80
273 van Haeringen WA, Gwakisa PS, Mikko S, Eythorsdottir E, Holm LE, Olsaker I,
274 Outteridge P & Andersson L (1999) Heterozygosity excess at the cattle DRB
275 locus revealed by large scale genotyping of two closely linked microsatellites.
276 *Animal Genetics* 30: 169-76
277
278