BRIEFINGS IN FUNCTIONAL GENOMICS. VOL 9. NO 3. 193–207 doi:10.1093/bfgp/elq002

# Fine-mapping and mutation analysis of *TRPM1*: a candidate gene for leopard complex (*LP*) spotting and congenital stationary night blindness in horses

*Rebecca R. Bellone, George Forsyth, Tosso Leeb, Sheila Archer, Snaevar Sigurdsson, Freyja Imsland, Evan Mauceli, Martina Engensteiner, Ernest Bailey, Lynne Sandmeyer, Bruce Grahn*, Kerstin Lindblad-Toh* and Claire M. Wade**

Advance Access publication date 29 March 2010

## Abstract

Leopard Complex spotting occurs in several breeds of horses and is caused by an incompletely dominant allele (*LP*). Homozygosity for *LP* is also associated with congenital stationary night blindness (CSNB) in Appaloosa horses. Previously, *LP* was mapped to a 6 cm region on ECA1 containing the candidate gene *TRPM1* (*Transient Receptor Potential Cation Channel, Subfamily M, Member 1*) and decreased expression of this gene, measured by qRT–PCR, was identified as the likely cause of both spotting and ocular phenotypes. This study describes investigations for a mutation causing or associated with the Leopard Complex and CSNB phenotype in horses. Re-sequencing of the gene and associated splice sites within the 105 624 bp genomic region of *TRPM1* led to the discovery of 18 SNPs. Most of

Corresponding author. Rebecca Bellone, Department of Biology, University of Tampa, 401 W. Kennedy Blvd. Box 3F, Tampa, FL 33611, USA. Tel: +1-813-257-3551; Fax: +1-813-258-7881; E-mail: rbellone@ut.edu

*These three authors contributed equally to this work.

**Rebecca R. Bellone** is an associate professor of Biology at the University of Tampa, USA. Her research is focused on understanding the molecular mechanisms of pigmentation variation in horses.

**George Forsyth** is a professor in Veterinary Biomedical Sciences at the University of Saskatchewan with interests in molecular genetics and physiology of ion channels.

**Tosso Leeb** is professor for veterinary genetics and animal breeding at the University of Berne. His main research interest is the elucidation of DNA variation underlying heritable traits in domestic animals.

**Sheila Archer** is an independent researcher. Her focus is quantitative inheritance through the analysis of phenotypic expression of coat color traits in horses.

**Snaevar Sigurdsson** is a post-doctoral fellow at the Broad Institute MIT and Harvard in the Vertebrate Genome Biology Group. His research is focused on genetics of complex diseases using the dog as a model animal.

**Freyja Imsland** is a PhD student at the Department of Medical Biochemistry and Microbiology at Uppsala University. Her thesis research is focused on the molecular genetics of coat colors in horses and molecular characterization of phenotypic trait loci in chickens.

**Evan Mauceli** is a Computational Biologist at the Broad Institute MIT and Harvard in the Vertebrate Genome Biology Group.

**Martina Engensteiner** is a veterinary. She has written her doctoral thesis about color genetics at the University of Berne.

**Ernest Bailey** is a professor at the University of Kentucky, Department of Veterinary Science. He conducts research on horses using genomics.

**Lynne Sandmeyer** is an associate professor of Ophthalmology at the Western College of Veterinary Medicine, University of Saskatchewan. She is a board certified member of the American College of Veterinary Ophthalmologists. Her interests include ocular pathology and investigation of inherited ocular disease in large and small animals.

**Bruce Grahn** is a professor in Small Animal Clinical Sciences and is Associate Dean Academic for the Western College of Veterinary Medicine. His research focuses on inherited ocular disorders especially developmental retinal disorders in Domestic animals, ocular pathology of animals and glaucoma shunt development.

**Kerstin Lindblad-Toh** is a professor in comparative genomics at Uppsala University Sweden and the Director of Vertebrate Genome Biology at Broad Institute of MIT and Harvard. Her research focuses on annotating the human genome through comparative analysis of vertebrate genomes as well as on the development of mapping tools for vertebrate model organism. She has a large group focusing on canine disease mapping.

**Claire M. Wade** is a Professor and Chair of Animal Genetics and Computational Biology at the University of Sydney in NSW Australia. Her research explores genetic variation in mammals and uses discovery of new variants to design genomic resources for companion animals.

the SNPs did not have a predictive value for the presence of *LP*. However, one SNP (ECA1:108,249,293 C>T) found within intron 11 had a strong ($P < 0.0005$), but not complete, association with *LP* and CSNB and thus is a good marker but unlikely to be causative. To further localize the association, 70 SNPs spanning over two Mb including the *TRPM1* gene were genotyped in 192 horses from three different breeds segregating for *LP*. A single 173 kb haplotype associated with *LP* and CSNB (ECA1: 108,197,355 - 108,370,150) was identified. Illumina sequencing of 300 kb surrounding this haplotype revealed 57 SNP variants. Based on their localization within expressed sequences or regions of high sequence conservation across mammals, six of these SNPs were considered to be the most likely candidate mutations. While the precise function of TRPM1 remains to be elucidated, this work solidifies its functional role in both pigmentation and night vision. Further, this work has identified several potential regulatory elements of the *TRPM1* gene that should be investigated further in this and other species.

***Keywords:*** *Leopard Complex spotting; congenital stationary night blindness; Transient Receptor Potential Cation Channel; Subfamily M, Member 1; fine mapping; Illumina DNA sequencing*



## INTRODUCTION

Horses (*Equus caballus*) are valued by breeders and enthusiasts for their beauty and variety of coat color and white spotting patterns. The genetic bases for several different white spotting patterning in horses have been reported, including those for Frame Overo, Sabino-1, Tobiano, Dominant White spotting and Gray [1–5]. The gene for Leopard Complex spotting (also referred to as Appaloosa spotting) has been mapped to horse chromosome 1 (ECA1) but the molecular basis has not been identified [6].

Leopard Complex spotting is found in several breeds of horses including Appaloosa, Knabstrupper, Noriker, Pony of the Americas, American Miniature and British Spotted Pony, among others. This spotting is characterized by several different patterns and associated traits. All of these patterns are characterized by patches of white in the coat that tend to be symmetrical and centred over the hips [7]. The patterns differ by the extent of white on the individual (Figure 1) and range from horses that display only a few white flecks on their rump to horses that are nearly all white, known as the 'fewspot' pattern.

Leopard Complex spotting was named after one of the spotting patterns 'leopard' (Figure 1D) in which characteristic pigmented oval spots (termed 'leopard spots') are observed in the white patterned area [7, 8]. In addition to the white patterning and leopard spots, Leopard Complex spotting is associated with four additional pigmentation traits (known as 'characteristics') that include striped hooves, readily visible unpigmented sclera around the eye, mottled pigmentation around the anus, genitalia and muzzle, and *LP*-specific roaning (Figure 2) [8].

A single incompletely dominant autosomal locus, *LP*, controls for the presence or absence of these pigmentation patterns and their associated characteristics. This was established by following the breeding records of hundreds of horses in the Appaloosa breed as well as other breeds that segregate for *LP* [7, 9]. Modifiers are thought to be responsible for the multiplicity of patterns associated with Leopard Complex spotting [7–9]. Horses that inherit the dominant allele $LP^{LP}$ will have the associated characteristics as well as display one of several different patterns of white (Figure 1). Homozygotes $LP^{LP}/LP^{LP}$ tend to have few to no leopard spots, whereas heterozygotes typically have leopard spots in their white patterned area [7, 10]. Similar phenotypes are known to occur in other species but none have been reported to display the myriad of patterning associated with *LP* [11–13].

Often white spotting in mammals is associated with other pleiotropic effects. In humans hypopigmentation of the skin is often associated with developmental defects of the eye. These disorders are commonly known as albinism [14]. In horses the only reported link between white spotting and eye disorders occurs with the Leopard Complex spotting. We have recently reported an association between homozygosity for *LP* in the Appaloosa and congenital stationary night blindness (CSNB) [15, 16].

*LP* was linkage mapped to a 6 cM region on ECA1 between microsatellite markers *ASB8* and *1CA043* [6]. Included in this interval is the positional and functional candidate gene encoding the Transient Receptor Potential Cation Channel, Subfamily M, Member 1 (TRPM1) [17]. TRPM1 is a member of the transient receptor potential (TRP) channel family, a family of proteins thought

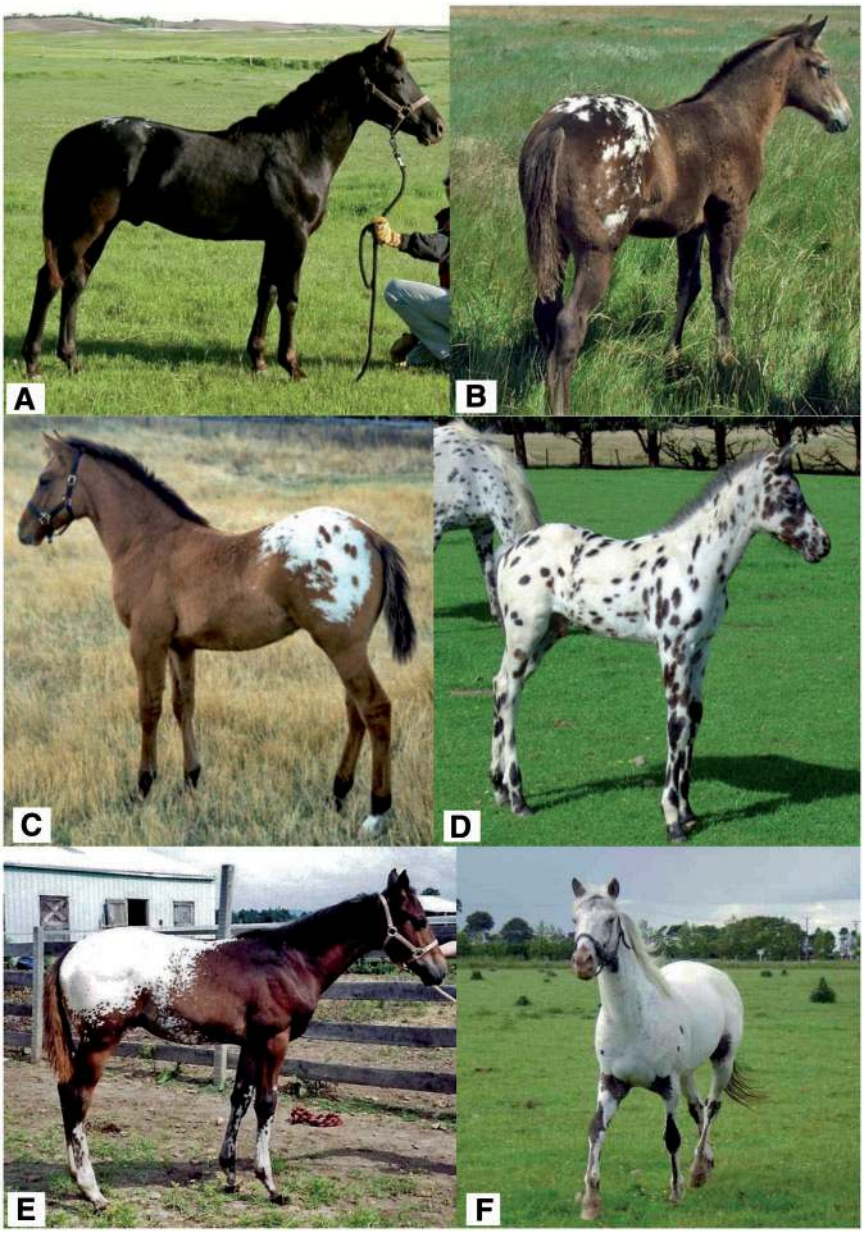


**Figure 1:** Leopard Complex spotting patterns. *LP* is characterized by several patterns which differ by the amount of white and the presence of leopard spots. (**A**) Horse displaying a few specks of white spotting on the rump. (**B**) Horse displaying the 'Lace Blanket' pattern, a small amount of white spotting is detected on the rump of the animal. (**C**) The 'Spotted Blanket' pattern, white is centred over the rump and can extend from the croup to the withers. Pigmented spots are present in the white patterned area. (**D**) The 'Leopard' pattern, white extends almost over the entire body and pigment spots are present in the white areas. (**E**) The 'Snowcap Blanket' pattern, white can extend from the croup to the withers but with few to no pigmented spots in the white patterned area. (**F**) The 'Few Spot' pattern, white extends almost over the entire body with few to no pigmented spots. All of these patterns are caused by a single dominant allele at the *LP* locus. E and F represent the homozygous genotypes, C and D are typical of heterozygotes and often the genotypes of horses similar to (A and B) cannot be determined from phenotype alone.

to be important in cellular and somatosensory perception [18]. Recently, qRT–PCR analyses showed that *TRPM1* mRNA expression is significantly down-regulated in both the skin of homozygotes ($LP^{LP}/LP^{LP}$) and the retina of CSNB-affected Appaloosas, whereas four other linked genes had unaltered mRNA expression [16]. This evidence suggests that the differential expression of *TRPM1* may

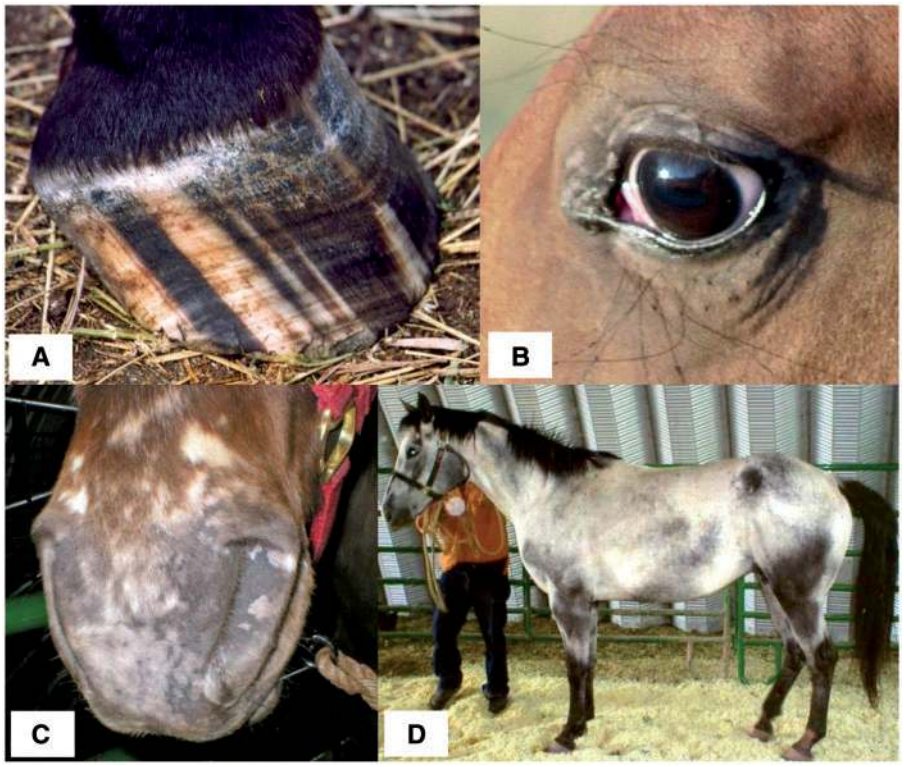


**Figure 2:** Leopard Complex associated characteristics. The characteristics associated with LP spotting include (**A**) striped hooves, (**B**) readily visible white sclera, (**C**) mottled skin and (**D**) varnish roaning. Mottled skin can occur around the anus, genitalia and as shown here around the eyes and muzzle (**B** and **C**). The roaning that occurs with the LP is different than that detected in other white spotting patterns and is termed 'varnish roan' in which there are white hairs interspersed with pigmented hairs but pigment (varnish marks) is retained on the bony surfaces, as seen here in (**D**) on the hip bone and face.

interfere with signaling for survival, differentiation and/or function of the melanocyte, in addition to the ON bipolar cell in the retinal rod pathway.

The causative mutation for *LP and* CSNB remains unknown although it is highly associated with the expression or stability of *TRPM1 mRNA*. In this study, we describe the mutation detection over the *LP* and CSNB locus. Within the previously linked interval we focused our efforts in the vicinity of *TRPM1*. First, the proximal promoter and coding regions (based on comparisons to human and mouse sequences) were sequenced. When no mutations explaining the expressional difference were detected, we refined the linked interval by fine mapping and identified potentially functional mutations by targeted sequencing.

## MATERIALS AND METHODS

### Horses

Horses from three different breeds known to segregate for *LP* were used in this study namely Appaloosa, Noriker and Knabstrupper. Horses from other breeds (Thoroughbred, American Quarter Horse and Paso Fino) not segregating for *LP* were used as controls. Thoroughbred samples were chosen from DNA previously collected and archived at the MH Gluck Equine Research Center. To identify polymorphisms within the coding regions DNA sequences from one solid horse ($LP^{lp}/LP^{lp}$) and one homozygous Appaloosa ($LP^{LP}/LP^{LP}$) were compared. To analyze the association of the intronic SNP identified with *LP* and CSNB, a total of 392 horses were tested: 146 Appaloosas, 34 Knabstruppers, 177 Norikers, 3 Quarter Horses and 32 Thoroughbreds. To fine map *LP* and CSNB, 192 horses were used: Appaloosa ($N=124$), Knabstrupper ($N=29$), Noriker ($N=36$) and Quarter Horse ($N=3$). For Solexa sequencing DNA from two $LP^{LP}/LP^{LP}$ homozygotes (1 Knabstrupper and 1 Appaloosa*)* was compared to the reference genome sequence and four other ($LP^{lp}/LP^{lp}$) controls (one Quarter Horse, one Icelandic Pony, one Lipizzaner and one Arabian).

## *LP* genotype determination

*LP* genotype was determined by breeding records and *LP* spotting pattern phenotype as documented by photograph or visual inspection. Specifically, horses were included in the $LP^{LP}/LP^{LP}$ group if they had an obvious coat pattern without leopard spots, i.e. 'fewspot' or 'snowcap blanket' (Figure 1E and F) [7, 10]. Horses in the $LP^{LP}/LP^{lp}$ group all displayed white patterning with leopard spots ('leopard' or 'spotted blanket') and/or had breeding records consistent with heterozygosity (only one *LP* carrying parent) (Figure 1C and D). Horses were included in the solid ($LP^{lp}/LP^{lp}$) group if they did not posses one of the leopard complex patterns and also did not show any of the other traits associated with the *LP*: striped hooves, white sclera and mottled skin (Figure 2).

## Ophthalmic examinations

Thirty of these horses used in both the SNP association and fine mapping studies were categorized by ocular examination as previously described [16]. Twenty-five of these horses were Appaloosas, two were Knabstruppers and three were American Quarter Horse controls. An a-wave-dominated ERG or 'negative ERG' was considered diagnostic of CSNB [15]. Fourteen horses had a 'negative ERG' and thus were diagnosed with CSNB and 16 controls (6 $LP^{LP}/LP^{lp}$ and 10 $LP^{lp}/LP^{lp}$) had normal scotopic and phototopic electroretinograms.

## DNA extraction

In addition to the previously archived Thoroughbred samples, whole blood or mane hair was collected by one of the authors or submitted by owners for this study. DNA from blood samples was extracted either using the Puregene whole-blood extraction kit (Qiagen Inc., Valencia, CA, USA) or Nucleon Bacc2 kit (GE Healthcare Bio-Sciences Corp., Piscataway, NJ, USA) according to the manufacturer's protocol. Hair samples were processed using 5–7 hair bulbs according to the method described [19]. Briefly, the hair bulbs were placed in a 100 µl lysis solution of FastStart Taq Polymerase PCR buffer (Roche, Mannheim, Germany), 2.5 mM $MgCl_2$ (Roche), 0.5% Tween 20 (JT Baker, Phillipsburg, NJ, USA) and 0.01 mg proteinase K (Sigma-Aldrich, St Louis, MO, USA) and incubated at 60°C for 45 min, followed by 95°C for 45 min to deactivate the proteinase K.

## DNA sequencing: putative regulatory regions and exons

Primers to amplify and sequence putative regulatory regions, exons and flanking introns of *TRPM1* were designed based on the publically available equine genome sequence. Specifically, the equine proximal promoter sequence was identified from the equine Trace Archive by performing BLAST searches using the human and mouse *TRPM1* proximal promoter sequences [20–22]. *Homo sapiens* miRNA within the *TRPM1* gene (Accession: MI0000287 ID: hsa-mir-211) was used to identify homologous miRNA sequence in horse [22–24]. Homo sapiens *TRPM1* transcript (Refseq: NM 002420, Ensembl transcript ID ENST00000256552) was used to identify equine homologous exons and flanking intron sequences by BLAST or BLAT searches [22, 25]. Exons 3–27 were confirmed by cDNA sequencing from retinal RNA. Primers were designed for the homologous horse sequences using Primer 3 software [26].

DNA from one solid non-Appaloosa ($LP^{lp}/LP^{lp}$) and one homozygous Appaloosa ($LP^{LP}/LP^{LP}$) was amplified using standard PCR procedures. Primer sequences, annealing temperature and product size can be found in Supplementary Table S1.

Amplicons were gel purified using the QIAquick gel extraction kit (Qiagen Sciences, MD, USA) and subsequently sequenced in both directions using BigDye® Terminator v1.1 and ABI 310 Genetic Analyzer (Applied Biosystems, Foster City, CA USA) or sent to the Core Sequencing Facility at the Plant Biotechnology Institute of the National Research Council, SK, Canada. Sequences were analyzed by aligning the $LP^{LP}/LP^{LP}$ and $LP^{lp}/LP^{lp}$ sequence data using ContigExpress from the Vector NTI Advance 10.3 software package (Invitrogen Corporation, Carlsbad, CA, USA). SNPs were identified by comparing these sequences with that of the EquCab2 genome assembly. Four SNPs detected in the $LP^{LP}/LP^{LP}$ sample when compared to the reference genome sequence were verified in a panel of 10 other $LP^{LP}/LP^{LP}$ horses.

## SNP association

A SNP detected from the $LP^{LP}/LP^{LP}$ sample in introns 11–12 (ECA1:108,249,293 C>T) showed complete association in the panel of 10 horses tested. The association of this SNP with *LP* and CSNB was further tested by either direct sequencing or by a *BsmFI* PCR-RFLP. Specifically PCR template for digestion was amplified in 10 µl PCR

reactions using 1× PCR buffer with 2.0 mM $MgCl_2$, 100 μM of each dNTP, 1 μl genomic DNA from hair lysate, 0.1 U FastStartTaq DNA polymerase (Perkin Elmer) and 200 nM of each primer (forward: 5′-GACTGAGCGTATGTGCGTGT-3′ and reverse: 5′-CTCCTGTCTGAGTGGCTTC A-3′). PCR reactions were performed at an annealing temperature of 64°C for 35 cycles. PCR products were digested for 14 h at 65°C according to the manufacturer's protocol using 2 U *BsmFI* (New England BioLabs Inc., Ipswich, MA, USA). Digested PCR products were analyzed on a 1.5% agarose gel stained with EtBr and illuminated by ultraviolet light. This RFLP was designed with an internal cut site that generated a 529 bp and 100 bp product. This enzyme also recognized the ECA1:108,249,293:C allele generating a product that was 362 bp and one that was 167 bp. A total of 357 horses from three different breeds known to have Leopard Complex spotting were tested. As a control group, 32 Thoroughbreds, not expected to have the *LP* mutation, were also tested. SNP genotypes were analyzed for the association with *LP* genotype by chi-squared analysis.

## Fine mapping

Seventy SNPs spanning over 2 Mb (ECA1:107, 194,138-109,299,508) and encompassing the candidate gene *TRPM1* were used to fine map *LP* and CSNB. These SNPs, previously identified in the 2007 (EquCab2) assembly (http://www.ensembl.org/Equus_caballus/index.html) of the horse genome, were genotyped in all 192 horses by the Sequenom Mass Spectrometry platform using the iPlex system [27]. The MassARRAY Designer software was used to design both PCR and MassEXTEND primers for all mutiplexed assays. The 70 SNPs were multiplexed in 35 plexes. All SNPs tested, PCR primer sequences and extension temperatures are listed in Supplementary Table S2. SNP genotypes were analysed for association with the *LP* genotype by the chi-squared analysis by breed. Then to maximize haplotype association, final analysis was performed with homozygous samples only ($LP^{LP}/LP^{LP}$ verses $LP^{lp}/LP^{lp}$) in all breeds tested using Haploview [28].

## Mutation screening of fine-mapped region

A 300 kb region of highest association was re-sequenced in two homozygous horses ($LP^{LP}/LP^{LP}$) (1 Knabstrupper and 1 Appaloosa) and four controls from different breeds ($LP^{lp}/LP^{lp}$) using hybrid capture (NimbleGen sequence Capture Arrays) followed by Illumina genome sequencing [29, 30]. In order to identify polymorphisms of interest, the short sequence reads (35 bp) were assembled and aligned with the horse reference genome sequence using MAQ alignment software [31]. This program uses a quality scoring matrix to assemble short reads to derive SNP calls based on a consensus diploid reference genome. Under a more stringent analysis, Spines was used to align all of the sequenced reads and a minimum coverage of 5-fold was required to call a polymorphic base (http://www.broadinstitute.org/science/programs/genome-biology/spines).

Any polymorphism identified in the two case samples ($LP^{LP}/LP^{LP}$) but not in the control ($LP^{lp}/LP^{lp}$) or the reference genome sequence (derived from a grey $LP^{lp}/LP^{lp}$ Thoroughbred) was further evaluated to determine if the mutation was located within conserved or transcribed regions. Conserved intervals were identified using Siphy [32] and comparing the SNP regions across 29 different eutherian mammalian species. Siphy allows for the comparisons of many related sequenced genomes to identify functional elements in a reference genome. The potential of SNPs to be transcribed was determined by comparing the location of individual SNPs with RNA sequence coverage data [33].

# RESULTS

## *TRPM1* targeted sequencing

Initial DNA sequencing was performed on coding sequence, flanking intron sequence (FIS) and putative regulatory regions of *TRPM1* including the proximal promoter, a putative internal miRNA gene, 5′- and 3′-UTRs and ∼0.5 kb upstream and 1.5 kb downstream from the gene. Thirty-six regions of *TRPM1,* including 15 459 bp of the 105 000 bp genomic region, were sequenced for two horses, one homozygous for the dominant allele ($LP^{LP}/LP^{LP}$) and another homozygous for the recessive allele ($LP^{lp}/LP^{lp}$) (Supplementary Table S1). Eighteen polymorphisms were found in the two horses sequenced, which were not present in the EquCab2 reference sequence. Of these, six SNPs occurred only in the solid non-appaloosa sample ($LP^{lp}/LP^{lp}$), seven occurred in both the $LP^{LP}/LP^{LP}$ and $LP^{lp}/LP^{lp}$ sequence and five SNPs were detected only in the $LP^{LP}/LP^{LP}$ sample but not the EquCab2

**Table 1:** *TRPM1* sequencing data

| Region Sequenced | Sequencing data coordinates | Confirmed cDNA sequence coordinates | SNP coordinates |
|---|---|---|---|
| 5′-flanking | 108,300,990:108,300,472 | | *lp*/*lp* 108,300,646 C>T |
| PP, exon 1, and FIS | 108,300,483:108,299,461 | | *lp*/*lp* 108,299,846 C>T (E) |
| | | | *lp*/*lp* 108,299,798 A>C (I) |
| Exon 2 and FIS | 108,274,240:108,273,867 | | ND |
| Exon 3 and FIS | 108,269,402:108,268,954 | 108,269,274:108,269,117[a] | ND |
| Exon 4 and FIS | 108,267,983:108,267,457 | 108,267,826:108,267,613 | *lp*/*lp* 108,267,614 G>A (E) |
| | | | ***LP*/*LP* 108,267,503 A>G (I)** |
| Exon 5 and FIS | 108,267,135:108,266,776 | 108,266,946:108,266,822 | *LP*/*LP* & *lp*/*lp* 108,267,115 A>G (I) |
| | | | *LP*/*LP* & *lp*/*lp* 108,267,097.108,267,098ins G (I) |
| | | | *LP*/*LP* & *lp*/*lp* 108,267,086.108,267,087insA (I) |
| Exon 6 and FIS | 108,266,295:108,265,631 | 108,266,007:108,265,836 | ND |
| Intron 6 with putative miRNA | 108,264,968:108,264,616 | | ND |
| Exon 7 and FIS | 108,263,668:108,263,069 | 108,263,429:108,263,255 | ND |
| Exon 8 and FIS | 108,263,022:108,262,508 | 108,262,818:108,262,695 | ND |
| Exon 9 and FIS | 108,262,143:108,261,661 | 108,261,915:108,261,843 | ND |
| Exon 10 and FIS | 108,261,190:108,260,663 | 108,260,983:108,260,883 | *LP*/*LP* & *lp*/*lp* 108,261,118 G>A (I) |
| Exon 11 and FIS | 108,249,623:108,249,098 | 108,249,484:108,249,311 | ***LP*/*LP* 108,249,293 C>T (I)** |
| | | | *LP*/*LP* & *lp*/*lp* 108,249,589.108,249,590delC (I) |
| Exon 12 and FIS | 108,248,341:108,247,813 | 108,248,098:108,247,964 | *lp*/*lp* 108,248,113 G>C (I) |
| | | | *lp*/*lp* 108,247,958 G>A (I) |
| | | | *LP*/*LP* & *lp*/*lp* 108,247,842 A>G (I) |
| Exon 13 and FIS | 108,247,322:108,246,917 | 108,247,166:108,247,116 | ***LP*/*LP* 108,246,967 C>T (I)** |
| | | | ***LP*/*LP* 108,247,024 C>T (I)** |
| Exon 14 and FIS | 108,244,813:108,244,522 | 108,244,727:108,244,587 | ND |
| Exon 15 and FIS | 108,243,754:108,243,533 | 108,243,697:108,243,668 | ND |
| Exon 16 and FIS | 108,239,337:108,239,013 | 108,239,294:108,239,002 | ND |
| Exon 17 and FIS | 108,236,872:108,236,676 | 108,236,847:108,236,619 | ND |
| Exon 18 and FIS | 108,229,735:108,229,351 | 108,229,620:108,229,498 | ND |
| Exon 19 and FIS | 108,229,348:108,228,947 | 108,229,294:108,229,163 | ND |
| Exon 20 and FIS | 108,227,233:108,226,753 | 108,227,072:108,226,944 | ND |
| Exon 21 and FIS | 108,224,519:108,224,176 | 108,224,461:108,224,210 | ND |
| Exon 22 and FIS | 108,223,420:108,223,147 | 108223371:108223197 | ND |
| Exon 23 and FIS | 108,221,660:108,221,467 | [b] | ND |
| Exon 24 and FIS | 108,221,033:108,220,599 | 108,220,834:108,220,690 | ND |
| Exon 25 and FIS | 108,219,100:108,218,779 | 108,219,061:108,218,859 | ND |
| Exon 26 and FIS | 108,218,287:108,218,090 | 108,218,232:108,218,100 | ND |
| FIS, exon 27 and 3′UTR | 108,198,136:108,196,462 | 108,197,956:108,196,448[a] | ***LP*/*LP* 108,197,261 A>C (E)** |
| 3′-flanking | 108,196,659:108,195,366 | | *LP*/*LP* & *lp*/*lp* 108,195,673 A>G |

Region sequenced, ECA1 coordinates for sequencing data, equine-specific cDNA ECA1 sequence coordinates and discovered SNPs are shown. All coordinates and base calls presented are according to the second assembly of the horse genome (EquCab2). PP, proximal promoter; FIS, flanking intron sequence; E, exon, I, intron, ND, none detected.

[a]Entire exons are not represented as primers to amplify cDNA were designed in these exons.

[b]This exon is present in humans but was not detected in either horse retina or skin transcripts. SNPs identified only in the *LP*/*LP* sample sequenced are in larger font and in bold print.

reference sequence and the $LP^{lp}/LP^{lp}$ sample. Of the latter five SNPs, only one SNP (ECA1: 108,197,261 A>C) occurred in the coding sequence, while the remaining four occurred within introns (Table 1). No associated polymorphisms were detected in any of the potential regulatory regions investigated (Table 1).

## SNP association

The exonic SNP detected (ECA1: 108,197,261 A>C) is expected to cause a missense mutation changing the 1464th amino acid of the protein from a methionine to arginine (p.M1464R). To further investigate the distribution of this SNP, DNA samples from seven additional $LP^{LP}/LP^{LP}$ homozygotes were sequenced for this region. The C allele was present only in two additional homozygotes, excluding this mutation as causative and thus this SNP was not investigated further.

The four intronic SNPs detected in the $LP^{LP}/LP^{LP}$ sample were validated in a panel of 10 unrelated

horses of all three genotypes. Two SNPs (ECA1:108,246,967 C>T and ECA1:108,247,024 C>T), located in introns 13–14, were not completely associated with the LP genotype, in that one LP/LP sample was homozygous for the C alleles. Thus, these SNPs were excluded as causative. Two other SNPs (ECA1:108,249,293 C>T and ECA1:108,267,503 A>G) showed perfect concordance with the *LP* genotype in this panel. There was no evidence of either mammalian conservation or transcription for either allele, although SNP ECA1:108,249,293 lies within 20 bp of exon 11 and ECA1:108,267,503 within 100 bp of exon 4. Since ECA1:108,249,293 C>T lies closest to the exon–intron boundary and could potentially be important for splicing, it was analyzsed further for association in a larger panel of 357 horses. In this larger sample, three breeds (Appaloosa, Knabstrupper, Noriker) segregating for LP contributed 295 horses, horses phenotyped for CSNB by a-wave-dominated ERG contributed 30, and Thoroughbreds known not to segregate LP contributed 32 controls.

There was a perfect association with homozygosity of the SNP (ECA1:108,249,293 C>T) and CSNB ($\chi^2 = 30$, $P < 0.0005$). In addition there was a complete association for this SNP and *LP* genotype in the Appaloosa ($\chi^2 = 298$, $P << 0.0005$) and Knabstrupper horses ($\chi^2 = 68$, $P < 0.0005$). Among horses of the Noriker breed, a strong association was observed ($\chi^2 = 120$, $P << 0.0005$). However, this association was not complete, in that 34 horses not suspected to have an *LP* allele possessed at least one copy of the T allele and two horses suspected to be heterozygous for *LP* were homozygous for this SNP (Table 2).

**Table 2:** ECA1: 108,249,293 C>T SNP association analysis in four horse breeds

| Breed, *N* and $\chi^2$ | *LP* genotype/ disease status | Genotype for SNP = T/T | Genotype for SNP = T/C | Genotype for SNP = C/C |
|---|---|---|---|---|
| Appaloosa ($N = 146$) ($\chi^2 = 292$, $P << 0.0005$) | | | | |
| | $LP^{LP}/LP^{LP}$ | 64 | 0 | 0 |
| | $LP^{LP}/LP^{lp}$ | 0 | 59 | 0 |
| | $LP^{lp}/LP^{lp}$ | 0 | 0 | 23 |
| CSNB ($N = 30$) ($\chi^2 = 30$, $P < 0.0005$) | | | | |
| | CSNB | 14 | 0 | 0 |
| | Unaffected | 0 | 6 | 10 |
| Knabstrupper ($N = 34$) ($\chi^2 = 68$, $P < 0.0005$) | | | | |
| | $LP^{LP}/LP^{LP}$ | 14 | 0 | 0 |
| | $LP^{LP}/LP^{lp}$ | 0 | 19 | 0 |
| | $LP^{lp}/LP^{lp}$ | 0 | 0 | 1 |
| Noriker ($N = 177$) ($\chi^2 = 120$, $P << 0.0005$) | | | | |
| | $LP^{LP}/LP^{LP}$ | 2 | 0 | 0 |
| | $LP^{LP}/LP^{lp}$ | 2 | 59 | 0 |
| | $LP^{lp}/LP^{lp}$ | 5 | 29 | 80 |
| Thoroughbred ($N = 32$) | | | | |
| | $LP^{LP}/LP^{LP}$ | 0 | 0 | 0 |
| | $LP^{LP}/LP^{lp}$ | 0 | 0 | 0 |
| | $LP^{lp}/LP^{lp}$ | 0 | 0 | 32 |

$\chi^2$-Analysis was performed to determine the association of the 108,249,293 C>T SNP with CSNB in Appaloosas and with *LP* in Appaloosas, Knabstruppers and Norikers. Thoroughbred is a breed of horses not expected to have the *LP* allele and thus served as a control. This SNP showed perfect association with CSNB and *LP* in Appaloosas and Knabstruppers but was less associated with phenotype in the Noriker breed. *LP* genotype was predicted by breeding and photographic records; CSNB phenotype was determined by ERG as described in the article. *N* represents the number of horses tested in each breed.

## Interval refinement by fine mapping

A set of 70 SNPs, identified as part of the whole-genome sequencing in the horse [34], flanking *TRPM1* and spanning 2.1 Mb (ECA1:107,194,138-109,299,508), were analyzsed in 192 horses [Appaloosa: (125) including 13 with confirmed CSNB; Quarter Horse: (2); Knabstrupper (29); Noriker (36)] (Table 3). When comparing only homozygous patterns ($N = 64$, $LP^{LP}/LP^{LP}$) to solid ($N = 50$, $LP^{lp}/LP^{lp}$) across the three breeds 14 SNPs within the wider region demonstrated strong significance ($P_{raw} < 1.0 \times 10^{-06}$) and a single 173 kb *LP*-associated haplotype was identified (ECA1:108,197,355-108,370,150) ($\chi^2 = 153$, $P = 4.05 \times 10^{-35}$) (Table 3 and Figure 3A). The most associated SNP occurred at position ECA1:108,370,091 ($\chi^2 = 117$, $P = 2.3 \times 10^{-27}$) in a region upstream of the suspected protein-coding region for *TRPM1 (*Figure 3B). The second most associated SNP occurred in an intron of *TRPM1* (intron 11) ECA1:108,248,113 ($\chi^2 = 97$, $P = 6.2 \times 10^{-23}$) (Table 3, Figure 3). Both of these SNPs are within the observed associated haplotype.

Comparing only CSNB-phenotyped horses (cases/controls = 14/16), three SNPs were found to be significantly associated ($P < 1 \times 10^{-06}$). The strongest association occurred close together at positions ECA1:108,370,091 and ECA1: 108,370,150 ($\chi^2 = 28$, $P = 1.6 \times 10^{-07}$) (Table 4, Figure 3). The only annotated gene in this region is *TRPM1* (Figure 3B).

## Targeted re-sequencing

A region (EquCab2.0 chr1:108,199,741-108,500,045), encompassing the *LP* haplotype detected

**Table 3:** Association analysis of 70 SNPs on ECA1 and LP genotype among Appaloosa, Knabstrupper and Noriker Horses

| SNP number | SNP position on ECA1 | Associated allele | Ratio of major and minor alleles (Case, Control) | $\chi^2$ | *P* |
|---|---|---|---|---|---|
| 1 | 107,194,138 | C | 23:105, 3:95 | 12 | 0.00050 |
| 5 | 107,464,256 | A | 89:37, 50:44 | 7.0 | 0.0080 |
| 6 | 107,542,173 | G | 105:9, 71:17 | 5.8 | 0.016 |
| 8 | 107,701,604 | C | 128:0, 93:7 | 9.2 | 0.0024 |
| 12 | 107,796,380 | C | 122:2, 80:16 | 16 | 5.3E-05 |
| 14 | 107,815,200 | C | 128:0, 91:9 | 12 | 0.00050 |
| **16** | **107,931,910** | **A** | **124:2, 74:22** | **26** | **4.0E-07** |
| **17** | **107,965,305** | **A** | **121:7, 41:53** | **71** | **3.2E-17** |
| 18 | 107,965,422 | G | 123:5, 78:20 | 15 | 8.9E-05 |
| 20 | 108,074,144 | T | 126:0, 84:14 | 19 | 1.2E-05 |
| 21 | 108,078,468 | G | 83:45, 42:46 | 6.3 | 0.012 |
| 22 | 108,128,461 | C | 83:43, 39:55 | 13 | 0.0003 |
| **23** | **108,128,561** | **T** | **122:0, 65:25** | **38** | **5.7E-10** |
| **24** | **108,131,916** | **A** | **128:0, 69:29** | **43** | **4.3E-11** |
| **25** | **108,132,170** | **T** | **126:0, 62:26** | **42** | **7.5E-11** |
| **26** | **108,132,263** | **C** | **128:0, 64:32** | **50** | **1.7E-12** |
| **28** | **108,181,934** | **T** | **127:1, 80:20** | **25** | **6.4E-07** |
| 29 | 108,182,386 | C | 126:0, 92:8 | 11 | 0.0012 |
| **30** | **108,197,355** | **C** | **125:1, 66:26** | **37** | **1.2E-09** |
| **31** | **108,227,501** | **C** | **128:0, 81:17** | **24** | **9.6E-07** |
| **32** | **108,248,113** | **G** | **128:0, 42:56** | **97** | **6.2E-23** |
| 34 | 108,329,772 | G | 127:1, 92:8 | 7.7 | 0.0055 |
| 35 | 108,340,357 | A | 120:0, 77:15 | 21 | 4.5E-06 |
| **36** | **108,343,655** | **C** | **128:0, 81:19** | **27** | **2.6E-07** |
| **37** | **108,370,091** | **T** | **128:0, 31:61** | **1.2 E02** | **2.3E-27** |
| 39 | 108,507,271 | A | 128:0, 97:3 | 3.9 | 0.048 |
| 40 | 108,549,650 | C | 128:0, 93:5 | 6.7 | 0.0098 |
| 43 | 108,827,565 | G | 119:5, 73:21 | 17 | 3.6E-05 |
| 44 | 108,832,497 | G | 128:0, 88:10 | 14 | 0.00020 |
| 47 | 108,861,525 | G | 119:5, 74:20 | 16 | 7.6E-05 |
| 57 | 108,991,827 | G | 122:4, 86:12 | 6.8 | 0.0089 |
| 58 | 108,992,653 | C | 124:4, 80:16 | 12 | 0.00040 |
| 59 | 108,992,843 | T | 121:1, 82:8 | 8.3 | 0.0040 |
| **60** | **109,023,700** | **G** | **121:5, 59:33** | **38** | **8.7E-10** |
| **61** | **109,045,266** | **G** | **109:19, 52:44** | **26** | **3.3E-07** |

Horses homozygous for *LP* ($LP^{LP}/LP^{LP}$, cases) where compared to solid non-characteristic horses ($LP^{lp}/LP^{lp}$, controls). Only those SNPs with $P \leq 0.05$ are presented. SNPs with $P_{raw} < 1.0 \times 10^{-06}$ are represented in bold.

by fine mapping, was investigated by NimbleGen hybrid capture for target sequence enrichment followed by Illumina Genome Analyzer sequencing. Initially, the sequence for analysis required base coverage >5-fold for each horse sequenced. In all, 359 differences, compared to the reference genome sequence, were identified in the sequenced region among all six horses sequenced (two $LP^{LP}/LP^{LP}$ and four $LP^{lp}/LP^{lp}$). A short insertion of varying length was detected at ECA1:108,222,721; however, one of the control horses also had this insertion, therefore it is not causative. Forty-three SNPs were detected in both of the homozygous $LP^{LP}/LP^{LP}$ horses but not in any of the controls and were thus identified as potential candidate mutations. Four of these SNPs were identified in the previous *TRPM1* target sequencing described above and thus were excluded as causative mutations. Specifically ECA1: 108,261,118 G>A is located in a highly conserved region across 29 eutherian mammals. However, during our *TRPM1* flanking intron sequencing investigation, the A-allele was detected in both the spotted ($LP^{LP}/LP^{LP}$) and non-spotted ($LP^{lp}/LP^{lp}$) samples; therefore, it is unlikely that this SNP is causative (Table 1). Of the remaining 39 potential candidate mutations, 3 are found in or near regions that are either transcribed or highly conserved (ECA1:108,281,765, 108,497,990 and 108,497,669, respectively) (Table 5). Using less stringent criteria (sequencing coverage to identify a SNP

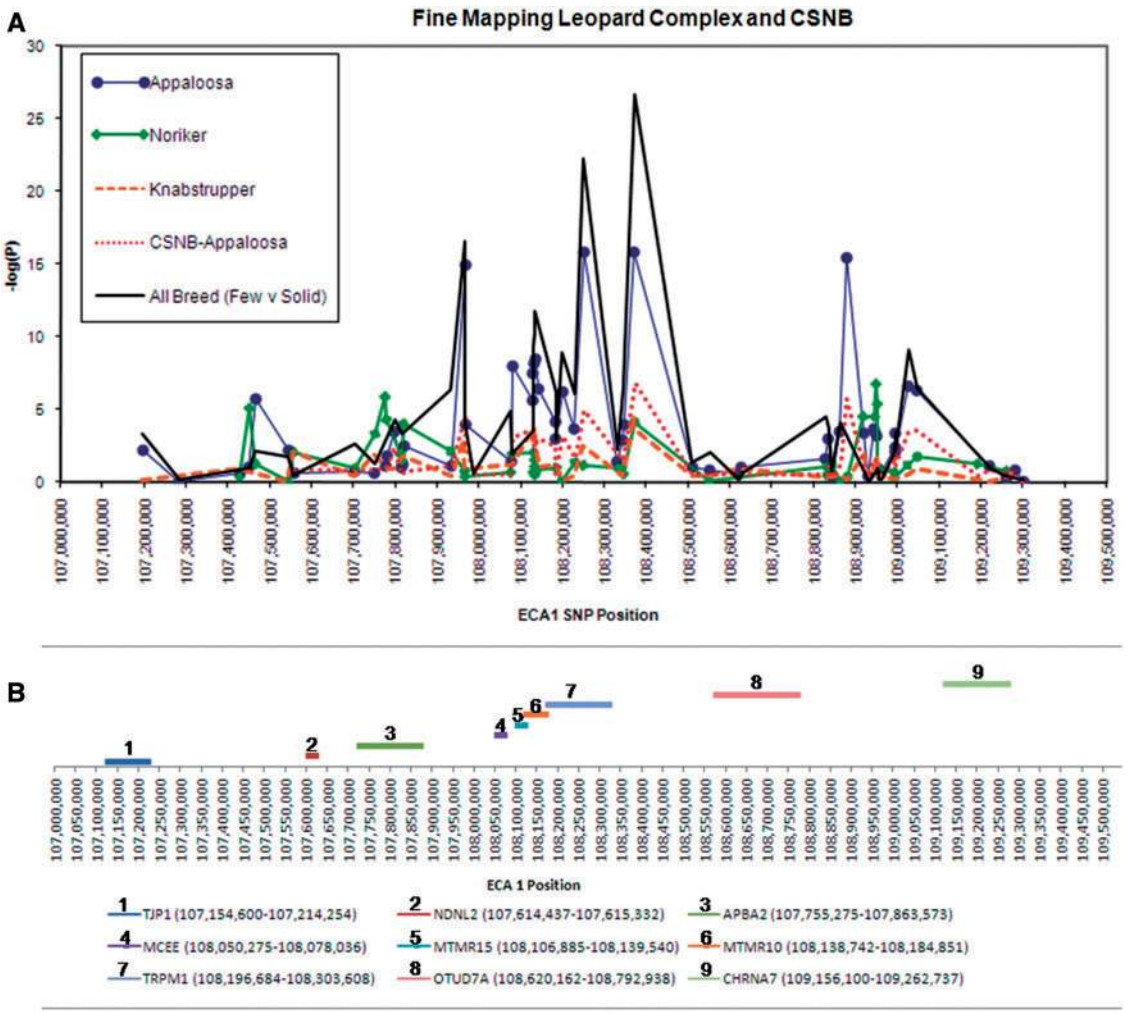


**Figure 3:** Fine mapping *LP* and CSNB. (**A**) Allelic association analysis by breed and combined for homozygous genotypes. (**B**) Ensemble Equcab2 gene predictions.

was 1 for the Appaloosa or Knabstrupper), 14 additional concordant SNPs were identified in the two $LP^{LP}/LP^{LP}$ horses sequenced, two of which are located in evolutionary conserved or transcribed regions (ECA1:108,288,853 and ECA1:108,489,901).

## DISCUSSION

Previous linkage and gene expression data implicated a *TRPM1* mutation as the cause for both *LP* and CSNB. DNA sequencing of the putative coding and regulatory regions identified one non-coding SNP (ECA1:108,249,293 C>T) that was highly associated with both CSNB and *LP*. However, this SNP did not show complete concordance with the *LP* genotype in the Noriker breed and we concluded that this SNP was not causative for *LP*. Nevertheless, because this SNP exhibits strong linkage disequilibrium with *LP*, it could be used as a genetic marker to test for zygosity for *LP* and CSNB for the Appaloosa and Knabstrupper horses, as many horses are born with low amounts of white patterning that do not allow a precise assignment of their genotype by their phenotype (Figure 1A and B).

Given the large difference in mRNA prevalence previously observed, it was hypothesized that a mutation in the proximal promoter causes this expression difference. Our results do not support this hypothesis; no mutation was detected in this region. However, sequence analysis provided verification that the equine proximal promoter is similar to both the described promoters for mouse and human having four potential MITF binding sites, a CAT box and a non-consensus TATA box (Figure 4) [20, 21].

Translation of the human full-length TRPM1 protein of 1533 amino acids was previously reported

**Table 4:** Association analysis of 70 SNPs on ECA1 and CSNB

| SNP number | SNP position on ECA1 | Associated allele | Ratio of major and minor alleles (case, control) | $\chi^2$ | *P* |
|---|---|---|---|---|---|
| 12 | 107,796,380 | C | 26:0, 19:7 | 8.1 | 0.0045 |
| 17 | 107,965,305 | A | 25:1, 11:15 | 18 | 2.6E-05 |
| 18 | 107,965,422 | G | 25:1, 16:10 | 9.3 | 0.0022 |
| 21 | 108,078,468 | G | 24:2, 13:13 | 11 | 0.00080 |
| 22 | 108,128,461 | C | 25:1, 13:13 | 14 | 0.00020 |
| 23 | 108,128,561 | T | 24:0, 17:7 | 8.2 | 0.0042 |
| 24 | 108,131,916 | A | 26:0, 20:8 | 8.7 | 0.0031 |
| 25 | 108,132,170 | T | 26:0, 18:6 | 7.4 | 0.0066 |
| 26 | 108,132,263 | C | 26:0, 18:8 | 9.5 | 0.0021 |
| 27 | 108,140,867 | T | 25:1, 15:11 | 11 | 0.0010 |
| 28 | 108,181,934 | T | 26:0, 20:8 | 8.7 | 0.0031 |
| 30 | 108,197,355 | C | 26:0, 18:10 | 11 | 0.00070 |
| 31 | 108,227,501 | C | 26:0, 23:5 | 5.1 | 0.024 |
| 32 | 108,248,113 | G | 26:0, 13:15 | 19 | 1.1E-05 |
| 35 | 108,340,357 | A | 24:0, 20:4 | 4.4 | 0.037 |
| 36 | 108,343,655 | C | 26:0, 22:6 | 6.3 | 0.012 |
| **37** | **108,370,091** | **T** | **26:0, 8:18** | **28** | **1.5E-07** |
| **38** | **108,370,150** | **C** | **26:0, 8:18** | **28** | **1.5E-07** |
| **48** | **108,878,149** | **C** | **23:3, 6:20** | **23** | **2.1E-06** |
| 51 | 108,942,855 | T | 26:0, 23:5 | 5.1 | 0.024 |
| 52 | 108,947,019 | C | 26:0, 24:4 | 4.0 | 0.045 |
| 57 | 108,991,827 | G | 26:0, 23:5 | 5.1 | 0.024 |
| 58 | 108,992,653 | C | 26:0, 20:6 | 6.8 | 0.0092 |
| 60 | 109,023,700 | G | 26:0, 16:10 | 12 | 0.00040 |
| 61 | 109,045,266 | G | 24:2, 11:13 | 13 | 0.00030 |

Horses that were diagnosed with CSNB (cases) were compared to those unaffected (controls) as diagnosed by ERG. Only those SNPs with $P \leq 0.05$ are presented. SNPs with $P_{raw} < 1.0 \times 10^{-06}$ are represented in bold.

to begin in exon 3 and end in exon 27 [35]. Equine gene predictions were not available when the *TRPM1* coding region sequencing was underway; therefore, the equine exonic structure (exons 3–27) as predicted by extrapolating from the human assembly (Refseq: NM 002420, Ensembl transcript ID ENST00000256552) was verified by cDNA sequencing from RNA isolated from the retina (Table 1). Exon 23 as denoted in human nomenclature (currently listed as the predicted exon 24 for horse by Ensembl transcript ID ENSECAT00000009109) was not detected in cDNA sequencing of either retina or skin. Therefore, sequencing the exons enabled the identification of deviations from human sequence and from the predicted gene structure. Further work is needed to characterize the upstream exons that have been recently identified in other species as described in refs [36, 37] and below.

Although a mutation in the proximal promoter or protein coding sequence that could explain the expressional difference detected was not identified, fine mapping reduced the candidate region from ~10 Mb to a considerably smaller region of 300 kb that included a single 173 kb associated haplotype (ECA1: 108,197,355-108,370,150). The region includes part of the coding sequence of *TRPM1* as well as a region upstream of the coding region for this gene. According to UCSC and Ensembl genome browsers, no functional genes have been annotated in this upstream region in the horse [38, 39]. Therefore, this region may contain important regulatory elements for *TRPM1* and thus should be further investigated for potential consensus regulatory elements. Furthermore, recently an additional 5′-exon was identified in the retina of mouse and rat (described as 1a) and melanocytes, retina and brain in humans (described as exon 0) [36, 37]. Surprisingly, this exon is located over 30 kb and 58 kb upstream of the described proximal promoter in mouse and human, respectively. This exon contains an in-frame translation start site and thus results in an alternative longer isoform of TRPM1. It is likely that an additional promoter exists that drives the expression of this alternative upstream exon and this should be investigated as the cause for CSNB and *LP* in horses. Human and mouse amino

**Table 5:** LP sequence variants detected by re-sequencing of a 300 kb region on ECA1

| ECA1 SNP coordinate | SNP | Candidate SNPs for further investigation |
|---|---|---|
| 108,224,578 | G>A | |
| 108,227,370 | C>T | |
| 108,227,480 | T>G | |
| 108,228,781 | C>A | |
| 108,230,099 | C>T | |
| 108,232,622 | T>A | |
| 108,240,955 | T>C | |
| 108,247,024 | C>T | |
| 108,247,619 | G>A | |
| 108,249,037 | G>A | |
| 108,249,293 | C>T | |
| 108,259,139 | C>T | |
| 108,259,387 | T>C | |
| **108,261,118** | **G>A** | **Conserved**[a] |
| 108,267,503 | A>G | |
| 108,275,971 | C>T | |
| **108,281,765** | **T>C** | **Potentially transcribed** |
| 108,284,434 | C>T | |
| **108,288,853**[b] | **C>T** | **Transcribed** |
| 108,291,026 | G>A | |
| 108,296,061 | G>T | |
| 108,297,544 | A>C | |
| 108,298,921[b] | A>T | |
| 108,305,829 | G>A | |
| 108,309,573[b] | T>C | |
| 108,332,922[b] | C>T | |
| **108,337,089** | **T>G** | **Potential regulatory region** |
| 108,343,133 | G>C | |
| 108,366,821 | A>G | |
| 108,370,777 | C>T | |
| 108,393,647 | G>C | |
| 108,401,809 | G>C | |
| 108,403,393 | C>T | |
| 108,403,505 | C>T | |
| 108,403,958 | C>T | |
| 108,404,206 | T>C | |
| 108,404,334 | T>C | |
| 108,404,665 | T>C | |
| 108,410,214 | G>C | |
| 108,410,277 | A>C | |
| 108,410,280[b] | T>C | |
| 108,410,456 | T>A | |
| 108,410,547 | A>G | |
| 108,410,780[b] | C>T | |
| 108,412,174 | C>T | |
| 108,416,351 | T>C | |
| 108,416,523 | A>G | |
| 108,430,452[b] | G>A | |
| 108,469,938[b] | G>A | |
| 108,470,018[b] | G>A | |
| 108,481,738[b] | G>A | |
| 108,485,437[b] | G>A | |
| **108,489,901**[b] | **G>A** | **Conserved** |
| **108,497,669** | **C>A** | **Conserved** |
| **108,497,990** | **C>T** | **Transcribed** |
| 108,498,342[b] | T>C | |
| 108,499,854[b] | A>T | |

57 SNPs detected from re-sequencing 300 kb of ECA1. Those SNPs observed only in the LP samples and not in any of the controls are presented.

[a]This SNP was previously identified in $LP^{lp}/LP^{lp}$ horse and thus is not causative as described in the 'Results' section.

[b]SNPs identified by less stringent analysis. Candidate SNPs for further investigation are bold and described in the discussion.


**Figure 4:** Horse *TRPM1* proximal promoter and exon 1 sequence. The region displayed includes potential MITF binding sites (3 E boxes and 1 M box highlighted, in bold face print and labeled) a CAAT motif, as well as a non-consensus TATA box (underlined and labeled). E Box numbering is according to the human proximal promoter (NT_010194 and Zhigi *et al.* 2004). Numbering is based on +1 at the transcription start site, which is marked with an arrow. Asterisks represents the SNP in exon 1 detected by sequencing the $LP^{lp}/LP^{lp}$ sample.

acid sequences were used to BLAT [25, 38] the horse genome and the homologous sequence for this exon was obtained (ECA1: 108,336,195-108,336,121). The expression of this exon in the horse should be verified by RNA seq or RT–PCR in retina and melanocytes. Our re-sequencing efforts did not identify any candidate mutations in this region. The closest candidate SNP (ECA1: 108,337,089) was 894 bp away and while this SNP is not in a region that has previously been shown to be transcribed or conserved, this SNP could be within a yet undescribed distal promoter and should be further investigated.

Re-sequencing over the 300 kb region of highest association using high stringency parameters (coverage of 5× or greater) identified three other polymorphisms in either highly conserved (1) or potentially transcribed regions (2). While using lower stringency (coverage of 1×) identified two additional polymorphisms (one transcribed and one conserved), which should also be investigated for the association with *LP* and CSNB in all breeds segregating for these phenotypes. Specifically, SNP ECA1:108,489,901 G>A and ECA1:108,497,669 C>A are located in the regions conserved across 29

mammalian species, ~200 kb upstream of *TRPM1,* not within another annotated gene [38], and therefore could contain important regulatory elements for *TRPM1*. One additional SNP ECA1:108,497, 990 C>T is also located upstream of the annotated *TRPM1* gene but may be transcribed as suggested by RNA sequencing data. If strong association between any of these three SNPs and either *LP* genotype or CSNB disease status is observed, then further investigation for potential gene regulatory functions would be warranted.

Two other identified SNPs (ECA1:108,281, 765 T>C and ECA1:108,288,853 C>T), based on human annotation (Ensembl transcript ID ENST00000256552), are located in the large first intron (~26 kb in length). However, RNA sequencing implies that 108,288,853 C>T is in a transcribed region while 108,281,765 T>C is located between two short transcribed regions [33]. This emphasizes the importance and necessity of species-specific gene annotation confirmation. In humans, at least five isoforms of TRPM1 exist [37], in addition to the longer isoform produced by exon 0 described above, another isoform known as 92 + TRPM1 results from an alternative exon with a start codon (exon 1′) located between exons 1 and 2. The horse RNA sequencing data [33] provide evidence that this exon may exist in the horse and therefore these SNPs should be investigated further for association and expression within TRPM1 isoforms in the skin and retina of the horse.

While the causative mutation for CSNB and *LP* remains unknown, we have identified a SNP (ECA1:108,249,293 C>T) within *TRPM1* that can be used for genotyping for CSNB and *LP* in Appaloosas until the causative mutation has been identified. We have also provided additional evidence that *TRPM1* is the genetic cause of both CSNB and *LP* and have identified six additional polymorphisms to examine for the association with *LP* and CSNB. This investigation is underway. Through our analysis we have discovered the need for further annotation and characterization of the equine *TRPM1* gene, specifically with respect to regulatory regions, characterization of alternative 5′-exons, as well as determining functional isoforms.

The precise function of TRPM1 remains to be fully elucidated. This protein belongs to the $Ca^{2+}$ transient receptor potential superfamily. $Ca^{2+}$ signaling and sensation have obvious roles in both cell migration and signaling. Thus through our studies on *LP* and CSNB it has become apparent that TRPM1 likely plays a role in both melanocyte and ON bipolar cell functioning. Several recent functional and comparative mapping studies in other species substantiate this argument. Work by Shen *et al.* suggested that TRPM1 is essential for synaptic function of the metabotropic glutamate receptor (MGluR6) pathway in ON bipolar cells as evidenced by the negative ERG (similar to the ERG observed in CSNB Appaloosas) in *Trpm1* knockout mice [40]. A similar form of CSNB has been diagnosed in humans, known as the Schubert–Bornshein type [41, 42]; however, to date, no pigmentation disorders have been reported with this form of CSNB in humans [43]. Mutations in an X-linked gene, *NYX,* have been shown to cause some of the CSNB cases in humans [44]. Very recently, in humans, several causative mutations for CSNB have been identified in TRPM1 [43, 45–46]. However, none of these have been correlated with gene expression and none are located in the regions of our candidate SNPs. Furthermore, several cases in humans were not explained by detected mutations [45]; thus, identifying the mutation that causes CSNB in Appaloosa horses may help to identify the genetic cause in yet unresolved Schubert–Bornshein types of human CSNB and may also help identify the molecular link between *NYX, TRPM1*, and other genes.

In mice, the appearance of white in the coat results from the absence of mature melanocytes caused by defects at various stages of melanocyte development. Many genes have been described that affect this development [47], but none have been shown to cause the many variations of spotting as seen with LP. Furthermore *TRPM1* has not been described as a coat color gene. Recent work, however, by Oancea *et al.* has demonstrated that *TRPM1* expression correlates with the melanin concentration suggesting a potential role for TRPM1 in the storage of melanin [37]. The functional question for LP spotting remains: is the white spotting caused by the lack of melanocytes, disruption in the function of melanocytes, or both. Transcriptional regulation and processing resulting in different isoforms for this and other *TRP* genes are thought to be responsible for functional variability [36]. Thus it is possible that different isoforms of TRPM1 may be involved in melanocyte migration while others are involved in melanin storage or transfer from melanocyte to keratinocyte or hair follicular cells. TRPM1 could be

important for calcium concentration gradients leading to melanocyte migration during development (as TRPM7 assists migrating fibroblasts) [48]. Failure of the calcium gradient oscillation might cause the absence of pigmentation seen in *LP* spotting patterns and associated characteristics. Furthermore, melanin storage deficiencies may explain why horses with *LP* roan with age. Identifying the genetic mutation affecting TRPM1 that causes *LP* and CSNB in the horse will both enrich our understanding of TRPM1 function and will help to explain yet unresolved forms of human CSNB and pigmentation disorders.

## SUPPLEMENTARY DATA

Supplementary data are available online http:\\bfgp.oxfordjournals.org.

### Key Points

- Utilization of the equine genome for regional SNP mapping and Illumina sequencing successfully identified six candidate SNPs for further investigation and provided strong evidence that *TRPM1* is the locus responsible for both CSNB and *LP*.
- LP spotting serves as an important model to gain further insight into melanocyte biology, retinal bipolar cell signaling and the genetic connection between these processes.
- Our Illumina data illustrate the importance of coverage consideration with these types of sequencing data; high coverage helps to eliminate not only erroneous base calls but also potential causative SNPs. SNPs from low coverage reads should be considered and investigated appropriately.
- Similar white spotting phenotypes occur in other species; however, no other species display the multiplicity of patterns controlled by a single locus. In addition, none of these phenotypes have been attributed to variants in the *TRPM1* gene.
- In humans, several different mutations have been shown to cause CSNB, none of which has been associated with a pigmentation disorder.
- *TRPM1* is expressed in multiple isoforms and these specific isoforms, as they relate to function, need to be characterized in the horse.

**Acknowledgements**

The authors wish to thank all of the horse owners who provided samples for DNA extraction We thank Elizabeth Sefton, Kimberly Soules, and Kaitlyn Duke for technical assistance as part of their Independent Study Courses at the University of Tampa as well as Karla Brown and Scott Lawson for their assistance in DNA extraction and Tara Biagi for technical assistance with fine mapping. We thank Dr James MacLeod and Steven Coleman for sharing their RNA sequencing data.

## FUNDING

This study was supported by the Heather Ryan and L. David Dubé Veterinary Health and Research Fund, the National Human Genome Research Institute and a Dana Faculty Development Grant from the University of Tampa.

## *References*

1. Metallinos D, Bowling AT, Rine J. A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung disease. *Mamm Genome* 1998;**9**:426–31.
2. Brooks SA, Bailey E. Exon skipping in the KIT gene causes a Sabino spotting pattern in horses. *Mamm Genome* 2005;**11**:893–9.
3. Brooks S, Lear TL, Adelson D, *et al*. A chromosome inversion near the KIT gene and the Tobiano spotting pattern in horses. *Cytogenet Genome Res* 2007;**119**:225–30.
4. Haase B, Brook SA, Schlumbaum A, *et al*. Allelic heterogeneity at the equine KIT locus in dominant white (W) horses. *PLoS Genet* 2007;**3**(11):e195.
5. Pielberg GR, Golovko A, Sundström E, *et al*. A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nat Genet* 2008;**40**:1004–9.
6. Terry RB, Archer S, Brooks S, *et al*. Assignment of the appaloosa coat colour gene (LP) to equine chromosome 1. *Anim Genet* 2004;**35**:134–7.
7. Sponenberg DP, Carr G, Simak E, *et al*. The inheritance of the leopard complex of spotting patterns in horses. *J Hered* 1990;**81**:323–31.
8. Sponenberg DP, Archer S, Bellone R. Patters of White with Symmetric White Patches: the Leopard Complex. In: Sponenberg DP. *Equine Color Genetics, 3rd edn*. Ames, IA: Iowa State University Press, 2009:110–20.
9. Miller RW. Appaloosa Coat Color Inheritance. *Dissertation*. Bozeman, Montana: Animal Science Department, Montana State University, 1965.
10. Lapp RA, Carr G. Applied appaloosa color genetics. *Appaloosa J* 1998;**52**:113–5.
11. Stephenson DA, Lee KH, Nagle DL, *et al*. Mouse rump-white mutation associated with an inversion of chromosome 5. *Mamm Genome* 1994;**5**:342–8.
12. Jordan SA, Jackson IJ. A late wave of melanoblast differentiation and rostrocaudal migration revelaed in patch and rump-white embryos. *Mech Dev* 2000;**92**(2):135–43.
13. Stritzel S, Wohlke A, Distl O. A role of the microphthalmia-associated transcription factor in congenital sensorineural deafness and eye pigmentation in Dalmatian dogs. *J Anim Breed Genet* 2009;**126**:59–62.
14. King RA, Hearing VJ, Creel DJ, *et al*. Albinism. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds). *The Metabolic and Molecular Bases of Inherited Disease*. 7th edn. New York: McGraw-Hill, 1995:4353–92.
15. Sandmeyer L, Breaux CB, Archer S, *et al*. Clinical and electroretinographic characteristics of congenital stationary night blindness in the Appaloosa and the association with the leopard complex. *Vet Ophthalmol* 2007;**10**:368–75.
16. Bellone RR, Brooks SA, Sandmeyer L, *et al*. Differential gene expression of TRPM1, the potential cause of congenital stationary night blindness and coat spotting patterns

(LP) in the Appaloosa horse (*Equus caballus*). *Genetics* 2008; **179**:1861–70.

17. Bellone R, Lear T, Adelson DL, *et al*. Comparative mapping of oculocutaneous albinism type II (OCA2), transient receptor potential cation channel, subfamilyM member 1 (TRPM1) and two equine microsatellites, ASB08 and 1CA43, among four equid species by fluorescence *in situ* hybridization. *Cytogenet Genome Res* 2006;**114**:93A.
18. Nilius B. TRP channels in disease. *Biochim Biophys Acta* 2007;**1772**:805–12.
19. Locke MM, Penedo MC, Bricker SJ, *et al*. Linkage of the grey coat colour locus to microsatellites on horse chromosome 25. *Anim Genet* 2002;**33**(5):329–37.
20. Hunter JJ, Shao J, Smutko JS, *et al*. Chromosomal localization and genomic characterization of the mouse melastatin gene (Mlsn1). *Genomics* 1998;**54**:116–23.
21. Zhiqi S, Soltani MH, Bhat KM, *et al*. Human melastatin 1 (TRPM1) is regulated by MITF and produces multiple polypeptide isoforms in melanocytes and melanoma. *Melanoma Res* 2004;**14**:509–16.
22. Altschul SF, Madden TL, Schäffer AA, *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;**25**:3389–402.
23. Lim LP, Glasner ME, Yekta S, *et al*. Vertebrate microRNA genes. *Science* 2003;**299**(5612):1540.
24. Griffiths-Jones S, Saini HK, van Dongen S, *et al*. miRBase: tools for microRNA genomics. *Nucleic Acids Res* 2008;**36**: D154–8.
25. Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res* 2002;**12**(4):656–64.
26. Rozen S, Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds). *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Totowa, NJ: Humana Press, 2000:365–86.
27. Gabriel S, Ziaugra L, Tabbaa D. SNP Genotyping using the sequenom MassARRAy iPLEX platform. *Curr Protoc Hum Genet* 2009;**60**:2.12.1–2.12.18.
28. Barrett JC, Fry B, Maller J, *et al*. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**(2):263–5.
29. Hodges E, Zuan Z, Balija V, *et al*. Genome-wide in situ exon capture for selective resequencing. *Nat Genet* 2007; **39**(12):1522–7.
30. Gnirke A, Melnikov A, Maguire J, *et al*. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 2009;**27**:82–189.
31. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 2008;**18**(11):1851–8.
32. Garber M, Guttman M, Clamp M, *et al*. Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* 2009;**25**:i54–i62.
33. Coleman SJ, Zeng Z, Wang K, *et al*. Equine Gene Structure Annotation by RNA Sequencing. *J Equine Vet Sci* 2009; **9**(5):319–20.
34. Wade CM, Giulotto E, Sigurdsson S, *et al*. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 2009;**326**:865–7.
35. Xu XZ, Moebius F, Gill DL, *et al*. Regulation of melastatin, a TRP-related protein, through interaction with a cytoplasmic isoform. *Proc Natl Acad Sci USA* 2001; **98**:10692–7.
36. Lis A, Wissenbach U, Philipp SE. Transcriptional regulation and processing increase the functional variability of TRPM channels. *Naunyn Schmiedebergs Arch Pharmacol* 2005;**371**: 315–24.
37. Oancea E, Vriens J, Brauchi S, *et al*. TRPM1 forms Ion channels associated with melanin content in melanocytes. *Sci Signal* 2009;**2**(70):ra21.
38. Kent WJ, Sugnet CW, Furey TS, *et al*. The human genome browser at UCSC. *Genome Res* 2002;**12**:996–1006.
39. Hubbard TJP, Aken BL, Ayling S, *et al*. Ensembl 2009. *Nucleic Acids Res* 2009;**37**:D690–7.
40. Shen Y, Heimel JA, Kamermans M, *et al*. A transient receptor potential-like channel mediates synaptic transmission in rod bipolar cells. *J Neurosci* 2009;**29**(19):6088–93.
41. Witzel DA, Smith EL, Wilson RD, *et al*. Congenital stationary night blindness: an animal model. *Invest Ophthalmol Vis Sci* 1978;**117**:788–93.
42. Schubert G, Bornshein H. Beitrag zur A lyse des menschlichen electroretinogram. *Ophthalmolgica* 1952;**123**: 396–413.
43. Li Z, Sergouniotis PI, Michaelides M, *et al*. Recessive mutatis of the gene TRPM1 abrogate ON biopolar cell function and cause complete congential stationary night blindness in humans. *Am J Hum Genet* 2009;**85**(5):711–9.
44. Bech-Hansen NT, Naylor MJ, Maybaum TA, *et al*. Mutations in NYX, encoding the leucine-rich proteoglycan nyctalopin, cause X-linked complete congenital stationary night blindness. *Nat Genet* 2000;**26**(3):319–23.
45. Audo I, Kohl S, Leroy BP, *et al*. TRPM1 is mutated in patients with autosomal-recessive complete congenital stationary night blindness. *Am J Hum Genet* 2009; **85**(5):720–9.
46. van Genderen MM, Bijveld MMC, Claassen YB, *et al*. Mutations in TRPM1 are a common cause of complete congenital stationary night blindness. *Am J Hum Genet* 2009;**85**(5):730–6.
47. Baxter LL, Hou L, Loftus SK, *et al*. Spotlight on spotted mice: a review of white spotting mouse mutants and associated human pigmentation disorders. *Pigment Cell Res* 2004; **17**:215–24.
48. Wei C, Wang X, Chen M, *et al*. Calcium flickers steer cell migration. *Nature* 2009;**457**(7231):901–5.