5-2023

Genetic investigation of spontaneous harlequin coat patterning in a family of Finnish Collies

Megan Patterson

Genetic investigation of spontaneous harlequin coat patterning in a family of Finnish Collies

Megan Patterson
Dr. Leigh Anne Clark
Clemson University
Department of Genetics and Biochemistry
GEN 4920 Honors Thesis
5 May 2023

Abstract

The merle coat pattern is a pigmentary phenotype of dogs characterized by a dilute background with black patches. Merle is caused by a SINE insertion in *PMEL17*, a pigmentation gene expressed in melanocytes. The mutation causes aberrant splicing of transcripts and production of an abnormal protein. *Harlequin* is a dominant modifier of merle that further dilutes the background to white. Harlequin Great Danes have a heterozygous mutation impairing the ubiquitin-proteasome system, suggesting that the inability to degrade aberrant PMEL17 results in melanocyte death. Harlequin is not a recognized coat pattern of the Collie; however, a harlequin phenotype spontaneously appeared in a family of Finnish Collies. Pedigree analysis revealed an inheritance pattern consistent with a dominant modifier of merle occurring de novo in the proband. To identify the mutation, we generated 30X whole genome resequencing data from a second generation harlequin Collie. We identified over two million heterozygous variants and filtered for unique coding variants against 1400 canine genomes aligned to the CanFam3 reference genome. The remaining 155 variants were manually inspected in IGV revealing 10 candidate missense mutations. Sanger sequencing in family members revealed that none of the identified variants segregated with the harlequin phenotype. Future efforts will utilize newly available reference genomes to identify additional coding variants. The identification of novel mutations that cause harlequin patterning in merle dogs will provide insight into genes and sequences critical for proper functioning of the ubiquitin-proteasome system.

Introduction

Merle is a coat color of dogs characterized by dilute background pigmentation with black patches. Merle is an autosomal, incompletely dominant phenotype [1]. *SILV* (also known as

Pmel17) is a locus known to be responsible for hair pigmentation through activity in melanosomes [2,3]. The protein is expressed exclusively in melanocytes and appears to be necessary to form a matrix of fibers to facilitate melanin synthesis in a melanocyte [4]. Analysis by Clark et al. 2006 revealed a short interspersed element (SINE) at the boundary between intron 10 and exon 11 of *SILV* that segregates with the merle phenotype. The SINE contains a cryptic splice acceptor site, which when used, results in part of the SINE being retained in the final mRNA transcript [5]. The incorrect splicing occurs at the final intron/exon boundary and the reading frame is maintained, allowing the PMEL protein to be synthesized with a 52 amino acid insertion [6]. Aberrant protein is tagged for degradation by the ubiquitin-proteasome system, resulting in an overall decrease of functional PMEL protein. However, variability in splicing of each transcript allows for some functional PMEL protein to be made in each melanocyte. Dilute pigmentation comes from haploinsufficiency of PMEL, as the fibril matrix is necessary for full pigmentation of a melanocyte, and a single wild-type allele is not sufficient for full pigmentation [6].

Harlequin coat patterning is distinguished by dark spots on a white background and is the result of two heterozygous mutations, one at the merle locus and one at the harlequin locus [7]. Harlequin has no phenotypic effect on coat color in non-merle dogs, but a single copy of the harlequin allele in a merle dog results in a further dilution of pigment to white. This is consistent with harlequin being a dominant modifier of merle. Clark et al. 2011 identified the harlequin mutation in Great Danes as a missense mutation resulting in an isoleucine replacing a valine in PSMB7. *PSMB7* encodes the beta-catalytic subunit of the proteasome and therefore is involved in recognition and degradation of poly-ubiquitinated proteins. This suggests that the increased demand for proteasome function caused by aberrant PMEL cannot be met in a melanocyte with

impaired proteasome function. Mutant PMEL may reach a toxic level and contribute to death of melanocytes, causing a complete lack of pigmentation. This mutation is only found in Great Danes, and harlequin is not a recognized pattern of any other breed [8].

A spontaneous harlequin mutation appeared in a family of Finnish Collies with no previous history of harlequin patterning. The proband, Mansikki, is shown in **Figure 1.** A pedigree was assembled to examine the pattern of inheritance of harlequin in the family (**Figure 2**). Here I present my investigation aimed at identifying the harlequin mutation in this lineage of Finnish Collies.

Figure 1. Finnish harlequin Collie

The proband, Mansikki, was born with a white background coat color with large black spots. She was the first harlequin dog in her lineage.

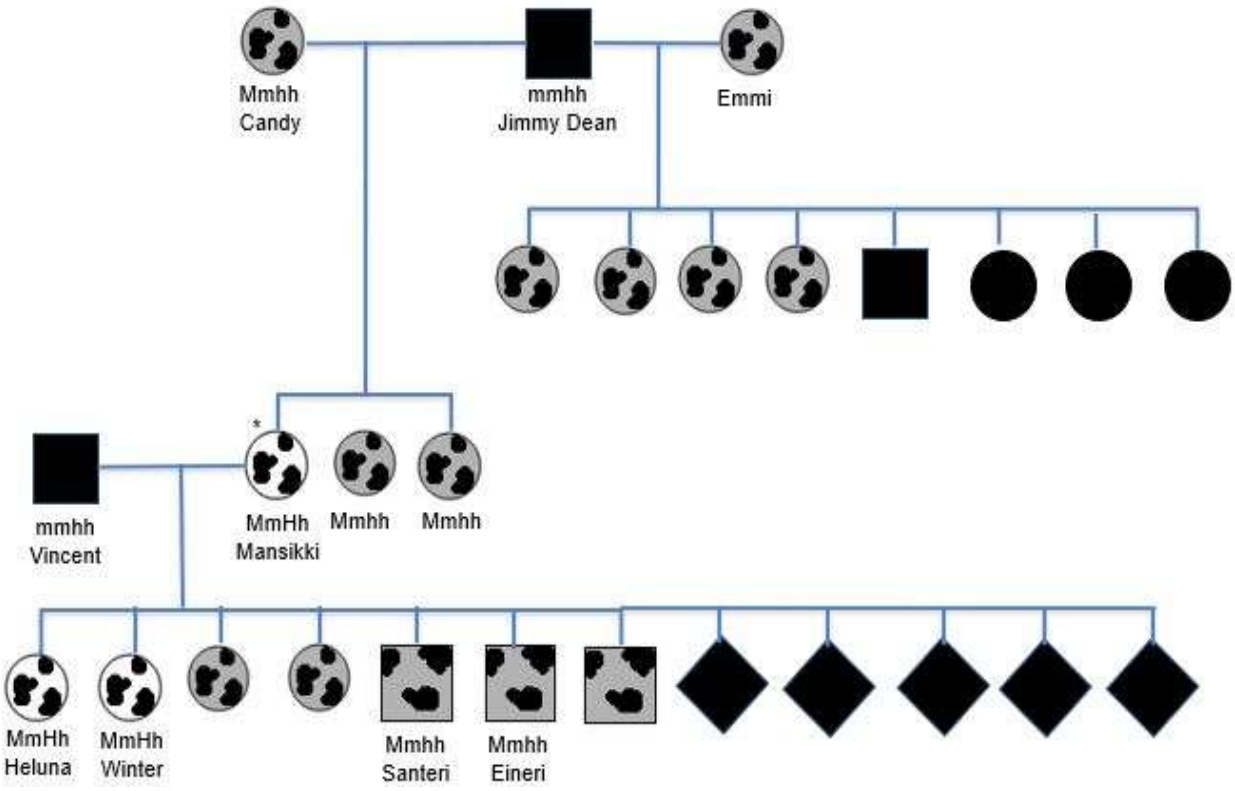

Figure 2. Finnish Collie pedigree

The proband, Mansikki, was born from a breeding between a tricolor sire and standard blue merle dam. A mating between Mansikki and a tricolor sire created two second generation harlequin, five merle, and five tricolor Collies.

Materials and Methods

Sample collection

The dog breeder submitted blood samples from the family of Finnish Collie,s and DNAs were isolated from white blood cells using standard protocols. DNA concentration was measured via a NanoDrop 1000 spectrophotometer, and samples were diluted to 50 ng/µL.

SNP array with five family members

DNA was prepared from the harlequin dam, the tricolor sire, one harlequin offspring, and two merle offspring. A SNP array was performed to generate genome-wide SNP profiles. Genotypes were used to identify the regions of the dam's genome inherited by the harlequin offspring, but not by the merle offspring.

Whole genome resequencing and variant filtering

DNA from a second generation harlequin puppy was used to generate high-coverage (30X) whole genome resequencing. The genome was aligned to CanFam3, and the resulting variant call file (VCF) was further filtered to narrow the list of candidate variants. Heterozygous variants were filtered against the 1399 canine genome reference panel with every previously identified mutation being removed from the VCF. The VCF was then filtered for coding mutations, leaving 155 unique, heterozygous, coding variants. The remaining variants were individually investigated using an integrated genome viewer to determine if the mutations were nonsynonymous. This left ten missense mutations.

A second round of filtering was done by adding back all unique heterozygous mutations and filtering by region to leave only variants in the critical regions. This left 518 variants within the critical intervals that were each individually investigated using the integrated genome viewer to determine if the variants were unique, heterozygous, and missense.

Sanger sequencing of family members

The 10 remaining variants were sanger sequenced in family members to determine if the variant segregated with the harlequin phenotype. Ten sets of primers were designed (**Table 1**) to amplify the genomic regions containing each variant.

Gene	Forward Primer	Reverse Primer
BRAT1	GGACACGAGACAAGCAGCAG	AGATTTGGTGGAGAGGTCAGG
SMG9	TGGAAGAAGTGAAGGTAGAATAGAT	TGGAACAAGTGGATACAGGGA
MAN1C1	ACTCACCATCACCACTTCCC	CACTTTATTTGACCTACTGTTCCC
DNAJC11	CAGACCCTGAGTGAACAATGG	GAAGTAAGTCCTGACCTGGCTG
ROGDI	TGTAAGTATGTTGCCCTGGAGAA	CCCACCTTATCCTTGAGTTGC
TMEM56	CATCACCATCCTTCAAAGAGAGT	ATGCTGCTACCATTGGCTGT
LPAR2	CCTACCTCTTCCTTATGTTCCAC	TGCTCAGATGTCACCTCCTTG
CCND2	ATAGGACCTTTTTCTCTCACTTCT	TTTGTGCTTCTTTCTACGGG
LSG1	GCTGTGTTATTACCTTGTGTTCC	CTTGATTCTACTTGATGAGACCTTC
NFASC	CCAGAGAGGGTCACATCCATT	GCACAGAGAGGAAGGCAACA

Table 1: Primer sets utilized to amplify 10 genomic regions

Primers were designed to have a melting temperature between 56-59℃ and not vary by more than approximately 1℃ within a pair. None of the primers have a risk of forming self dimers or hairpins, and pairs have a cross dimer not exceeding -4.5 kcal/mol. All primers are denoted as 5'-3'.

The PCR reaction was performed with 6.5µL of DreamTaq master mix, 4.4µL of sterile PCR water, 1µL of both the forward and reverse primer at a concentration of 10µM, and 50ng of DNA. Reactions were run on a thermal cycler using the following protocol: initial denaturation at 95℃ for 3 minutes; 5 cycles of 95℃ for 30 seconds, 58℃ for 30 seconds, and 72℃ for 30

seconds; followed by 30 cycles of 95℃ for 20 seconds, 56℃ for 30 seconds, and 72℃ for 30 seconds; with a final extension period of 72℃ for 10 minutes. PCR products were run on a 1.2% agarose gel and visualized using ultraviolet light to validate the amplification of the correct fragment. PCR products were purified by combining 1μL DNA with 4μL of a 1:8 dilution of EXOSAP-it solution and run in a thermal cycler at 37℃ for 1 hour, followed by a second incubation at 80℃ for 15 minutes. These products were mailed to EtonBio along with 1μL of forward primer for all except CCND2 and LSG1, which were sent with the reverse primer for Sanger sequencing. The sequencing results were viewed on Chromas to determine the genotype in four family members.

Results and Discussion

Pedigree analysis revealed an inheritance pattern consistent with a dominant modifier of merle occurring de novo in the proband. We feel confident that the sire of the proband was not a silent carrier of a harlequin mutation because he only produced one harlequin offspring through several matings with merle dams. The harlequin mutation was passed to two offspring, indicating the mutation occurred de novo in the germline of the proband, allowing it to be heritable. This suggests that the harlequin mutation is inherited in a dominant, heterozygous manner. A *de novo* mutation indicates that the mutation will be unique to this lineage and not previously reported, as no other Collies have been found with the same phenotype. Additionally, the mutation is likely to be coding due to the full penetrance of the phenotype. We believe the affected gene will be involved in the ubiquitin-proteasome system, although this is a very complex system and not all associated proteins have been identified. This investigation could uncover additional genes involved in the ubiquitin-proteasome system to gain a better understanding of the process.

Initial filtering efforts revealed 10 unique, heterozygous, missense mutations in the harlequin genome, shown in **Table 2.**

Gene	Mutation	Function
BRAT1	Gly332Asp	DNA damage pathway
SMG9	Pro9Leu	Nonsense-mediated mRNA decay
MAN1C1	Arg541Lys	N-glycan processing
DNAJC11	Val361Ile	Heat shock protein, cristae formation
ROGDI	Arg238Cys	Desmosomes and intermediate filaments
TMEM56	Arg145His	Lipid homeostasis
LPAR2	Leu202Pro	Calcium mobilization
CCND2	Stop299Cys	Cell cycle control
LSG1	Ala66Pro	GTPase
NFASC	Tyr1097Cys	Neuron growth and organization

Table 2: Final 10 heterozygous, unique, missense mutations

Ten mutations were identified as being unique, heterozygous, and missense in the harlequin dog. None of the identified genes have a known function in the ubiquitin-proteasome system.

None of the variants were particularly promising as none had previously described roles in the ubiquitin-proteasome system. However, each was analyzed through Sanger sequencing to determine if the variant segregated with the phenotype in the family. We expect that both harlequin dogs, Mansikki and Heluna, will be heterozygous for the harlequin mutation. Furthermore, because harlequin is a dominant modifier of merle, we know that the merle siblings, Santeri and Eineri, will be homozygous wild-type at the harlequin locus. Each variant

was investigated to confirm that the harlequin dogs were heterozygous while the merle dogs were homozygous wild-type. This was not the case for any of the 10 remaining variants, which narrowed the candidate list to zero variants. Our results suggest that one of the filtering steps may have removed the causal variant from the candidate list.

The use of annotation tracks to define coding regions may have resulted in the removal of potentially coding variants. To rectify this, we removed the coding filter and instead filtered by region to reveal only unique, heterozygous variants in the critical intervals identified through the SNP array, shown in **Table 3**.

Chr	**Start Position**	**End Position**	**Chr**	**Start Position**	**End Position**
1	82379071	98313613	**24**	19280317	47030908
5	29007870	36738109	**26**	6092231	39382709
7	51568622	60821231	**29**	37630167	43666041
8	66130863	76319553	**31**	4378156	37731982
10	57414624	69923852	**32**	20961027	29159102
17	39531645	47346811	**33**	9525550	15980689
19	3964287	5205552	**34**	10979881	13534261
23	16688219	17058739	**36**	3260911	32505072

Table 3: Critical regions likely harboring the harlequin mutation

A SNP array revealed the 16 regions of the genome shared by the two harlequin dogs, but not present in either merle offspring. The harlequin mutation is likely to be found in one of these critical regions.

Each of the 518 remaining variants were individually investigated to determine if they were heterozygous, coding, and unique to this lineage of dogs. None of these variants matched our hypothesis which again left us with zero candidate mutations.

From here, there are several options that will be taken to update the data available to find the mutation. One possible route is to focus on the entirety of each critical region without the VCF. This approach would eliminate the possibility of filtering the causal variant out of the candidate list due to mislabeling. While effective, this method would be time-consuming. A way to increase efficiency is to sequence additional genomes, from either the tricolor sire or a merle sibling. Additional data could help to narrow down the critical regions, and using full genome sequencing as opposed to SNP data would improve the accuracy of the borders of each region. Finally, the last option is to align Heluna's genome to a newer reference genome. More recent reference genomes may have more accurately characterized coding regions, allowing us to uncover additional coding variants. Identification of the mutation that causes harlequin in Finnish Collies will provide insight into genes critical for proper functioning of the ubiquitin-proteasome system.

Acknowledgements

I would like to thank my research advisor Dr. Leigh Anne Clark and past and present members of the Clark lab including Elizabeth Greif, Dr. Sarah Murphy Bell, and fellow undergraduate student Katherine Singleton for their assistance in this project. I would also like to thank Dr. Rooksana Noorai, who performed the variant filtering. Additionally, I would like to thank the breeder, Anna-Mari Lammi, for providing the samples. Finally, I would like to thank the Clemson Honors College for their help in funding this project.

References

1. Alan L. Mitchell, Dominant dilution and other color factors in collie dogs, *Journal of Heredity*, Volume 26, Issue 10, October 1935, Pages 425–430,
2. DUNN, L. C. 1936. Studies of multiple allelomorphic series in the house-mouse, I, Description of agouti and albino series in the allelomorphs. Journ. Genetics,*33*, p. 443–453.
3. Kwon BS, Halaban R, Ponnazhagan S, Kim K, Chintamaneni C, Bennett D, Pickard RT. Mouse silver mutation is caused by a single base insertion in the putative cytoplasmic domain of Pmel 17. Nucleic Acids Res. 1995 Jan 11;23(1):154-8. doi: 10.1093/nar/23.1.154. PMID: 7870580; PMCID: PMC306643.
4. Theos AC, Truschel ST, Raposo G, Marks MS. The Silver locus product Pmel17/gp100/Silv/ME20: controversial in name and in function. Pigment Cell Res. 2005 Oct;18(5):322-36. doi: 10.1111/j.1600-0749.2005.00269.x. PMID: 16162173; PMCID: PMC2788625.
5. Clark LA, Wahl JM, Rees CA, Murphy KE. Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. Proc Natl Acad Sci U S A. 2006 Jan 31;103(5):1376-81. doi: 10.1073/pnas.0506940103. Epub 2006 Jan 9. PMID: 16407134; PMCID: PMC1360527.
6. Clark LA, Wahl JM, Rees CA, Strain GM, Cargill EJ, Vanderlip SL, Murphy KE. Canine SINEs and their effects on phenotypes of the domestic dog. In: Gustafson JP, Taylor J, Stacey G, editors. Genomics of disease. New York: Springer; 2008. p. 79–88.
7. D. Phillip Sponenberg, Inheritance of the harlequin color in Great Dane dogs, *Journal of Heredity*, Volume 76, Issue 3, May 1985, Pages 224–225

8. Leigh Anne Clark, Kate L. Tsai, Alison N. Starr, Keri L. Nowend, Keith E. Murphy, A missense mutation in the 20S proteasome β2 subunit of Great Danes having harlequin coat patterning, Genomics, Volume 97, Issue 4, 2011, Pages 244-248