5-2023

Genetic investigation of spontaneous harlequin coat patterning in a family of Australian shepherds

Katherine Grace Singleton

Genetic investigation of spontaneous harlequin coat patterning in a family of Australian shepherds

Katherine Grace Singleton

Dr. Leigh Anne Clark

Clemson University

Department of Genetics and Biochemistry

GEN 4910 Honors Thesis

May 2023

ABSTRACT

In dogs, a white base coat with black patches is termed harlequin patterning. In the Great Dane, harlequin is caused by a mutation in the ubiquitin-proteasome system, a highly complex process through which damaged proteins are degraded[1]. All harlequins are also heterozygous for the *Merle* allele of *SILV*, which harbors a retrotransposon that leads to production of abnormal protein[2]. In melanocytes with impaired protein degradation, aberrant SILV is thought to cause cell death, resulting in the characteristic white base coat of a harlequin dog[1]. We identified a novel, spontaneous form of harlequin patterning in a family of Australian shepherds. To identify the genetic cause, we generated genome-wide SNP profiles and identified 32 chromosomal regions with genotype patterns consistent with the known mode of inheritance: heterozygous in the sire and harlequin offspring (n=4) and absent from the non-harlequin merles (n=2). We generated whole genome and skin transcriptome sequences from a harlequin dog to identify candidate causal variants. High quality, heterozygous variants were filtered against multibreed VCFs representing nearly 2000 canines to identify protein coding changes private to the harlequin Australian shepherds. Seven variants were evaluated through *in silico* programs for variant effect and sequenced to confirm inheritance pattern. We identified a candidate causal variant in *BRCA1*, an E3 ubiquitin ligase. To quantify the impact of the variant we will synthesize mutant SILV and evaluate the effectiveness of both mutant and wild type BRCA1 at performing ubiquitination.

INTRODUCTION

Merle is a coat pattern characterized by dark spots on a dilute background, typically black on grey. Merle is recognized as a standard breed coat pattern in breeds such as Great Danes, collies, Shetland sheepdogs, corgis, and Australian shepherd dogs. It is caused by a

retrotransposon insertion in the gene *SILV*, with an oligo(dT)-rich tail of variable length that determines the exact coat pattern. One of these variations, caused by the longest category of oligo(dT)-rich tail, is the harlequin coat pattern, characterized by dark patches on a white background[3]. However, in the Great Dane breed, *Harlequin* is inherited independently of *Merle* as an allele of *PSMB7*, the beta-2 catalytic subunit of the proteasome[1].

Figure 1. Ollie, a harlequin Australian shepherd dog and the proband of the study.

The proteasome is a large protein complex comprised of multiple subunits that serves to degrade flagged proteins in the cell. These proteins are marked for degradation by ubiquitin ligases, through a cascade involving E1, E2, and E3 proteins, with the E3 proteins performing the direct ubiquitination. Misfolded or aberrant proteins are ubiquitinated on lysine residues and sent to the proteasome, a process essential for cell viability[4]. In a system with impaired proteasome function, mutant Merle accumulates, causing melanocyte death[1].

The harlequin coat pattern is not recognized as breed standard in Australian shepherd dogs; however, we identified a harlequin Australian shepherd dog without an oligo(dT)-rich tail of the longest categorical length or the Great Dane mutation in *PSMB7*, suggesting that the coat pattern results from a different mechanism. We generated whole genome resequencing data and a skin transcriptome from the proband as well as SNP array data from the proband, sire, three

half-siblings, and the half-sibling's dam. SNP filtering began with searching for heterozygosity since, in the Great Dane, *Harlequin* is homozygous lethal[1]. Further filtering based on coding and splice site regions, non synonymous effect, uniqueness, and the critical intervals derived from the array data lead to seven candidate variants. Sanger sequencing and genotyping of the family revealed four *de novo* mutations out of our candidate seven. The correct variant has to be inherited since the proband's sire produced other harlequin offspring. We are now left with three candidate variants, the strongest of which is a p.Lys220Arg mutation in *BRCA1*, commonly known to be a tumor suppressor, but less commonly known to be an E3 ubiquitin ligase.

METHODS

Sample Collection

DNA samples isolated from whole blood or buccal swabs were available from previous studies. Additional samples were collected via buccal swabs and isolated using PureGene protocols. The family unit of interest consisted of a tri-colored sire, standard merle dam, their two harlequin progeny, and one of their standard merle progeny. The proband of the study was the harlequin offspring of the tri-colored sire and a different standard merle dam. Other dogs in the study included standard merle Australian shepherd dogs, collies, and Shetland sheepdogs.

High coverage whole genome resequencing data for the proband were generated using next generation sequencing technologies. Transcriptome sequencing was performed with RNA obtained from a pigmented skin punch biopsy. Whole genome SNP genotypes were generated using the Illumina CanineHD BeadChip for the defined family unit and the proband, as well as critical intervals which the sire, proband, and two harlequin progeny shared, but the dam and merle half-sibling did not.

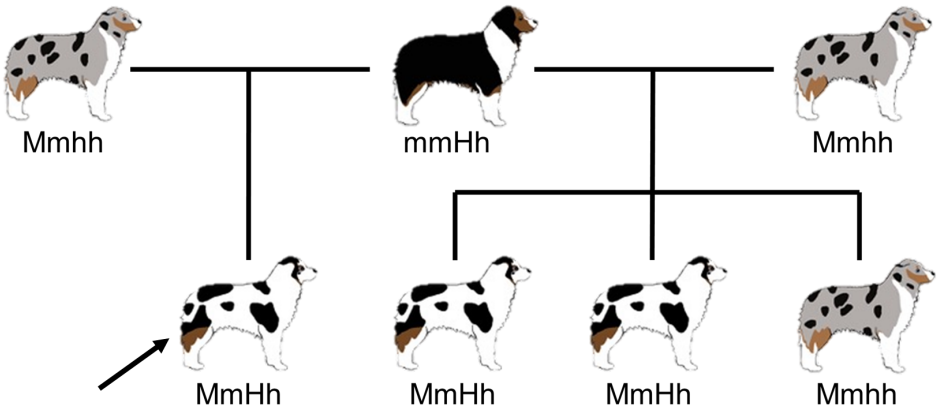

Figure 2. Pedigree showing the proband (indicated with arrow), dam, and sire's nuclear family. *M* indicates the *Merle* allele and *H* indicates the *Harlequin* allele.

Variant Filtering

The critical intervals of the whole genome and skin transcriptome were manually scanned using Integrated Genome Viewer 2.11.1, to look for variants heterozygous in both the genome and transcriptome. Variants were filtered from a variant call file generated from the whole genome using SNP Variation Suite Golden Helix 8.9.0. The VCF was crossed by chromosome with a file containing 722 canine genomes to filter for uniqueness. Remaining variants were then filtered for heterozygosity and for coding regions. Variants were manually sorted by critical interval and checked for synonymy. Remaining variants were checked in files containing 1399 and 628 canine whole genomes for uniqueness.

Genotyping

The family unit and proband were genotyped for each of the remaining seven variants. Primers were designed for each variant, and the DNA was PCR amplified using the primers. PCR products were spot checked with electrophoresis on an agarose gel, then ExoSAP purified and sent to EtonBio for Sanger sequencing. Chromatograms were analyzed using SnapGene to confirm the correct inheritance pattern.

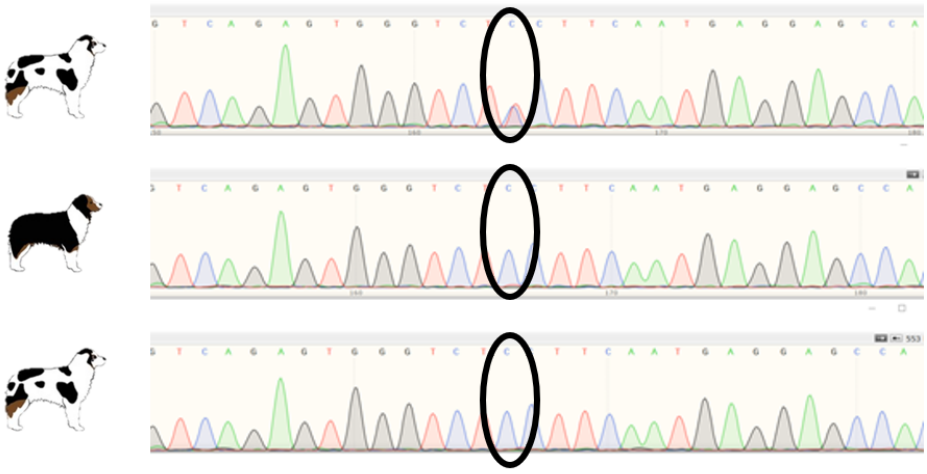

Figure 3. Chromatograms showing the *de novo* variant in *STON2*. The top is the harlequin proband, the middle is the tri-colored sire, and the bottom is a harlequin half-sibling. To fit our predicted inheritance pattern, all three dogs should be positive for the variant. The tri-colored sire and harlequin half-sibling are not.

A control population of merle Australian shepherd dogs, collies, and Shetland sheepdogs were genotyped for the remaining variants. Candidates were run through *in silico* programs to determine potential pathogenicity. Programs used were PANTHER, PolyPhen2, and SIFT.

RESULTS AND DISCUSSION

The variant filtering process resulted in seven candidate variants being identified in the following genes: *BRCA1, PHLDB2, SPPL2A, CELSR2, STON2, TMEM131L,* and *INTS2*. We came to the same seven variants both through the manual scanning and the bioinformatic filtering.

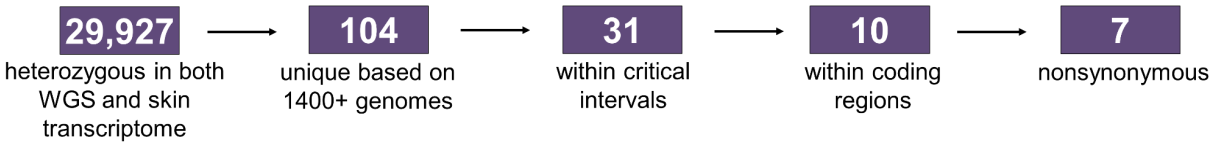

Figure 4. Flowchart showing the initial bioinformatic filtering of the variant call files.

All variants were evaluated through *in silico* programming, and their PolyPhen2 scores were 0.978, 1.000, 0.993, 0.951, 1.000, 0.000, and 0.999, respectively. Each variant was Sanger

sequenced in a population of 15 merle Australian shepherd dogs unrelated to the family. Since *Harlequin* is a dominant modifier of *Merle*, a merle dog cannot have a *Harlequin* allele and not express it. All variants were absent from the test population, supporting their uniqueness. Sanger sequencing of each variant in the proband and nuclear family revealed that the mutations in *PHLDB2, SPPL2A, CELSR2,* and *STON2* were *de novo* in the proband and therefore not the genetic cause of the harlequin coat pattern. Additionally, the *BRCA1* variant was sequenced in a population of 40 standard merle collies and Shetland sheepdogs. Collies and Shetland sheepdogs are the closest evolutionarily related breeds to Australian shepherds, making them ideal additional controls. All controls were negative for the *BRCA1* variant.

The three remaining candidate variants are in *BRCA1, TMEM131L,* and *INTS2*. *TMEM131L* codes for transmembrane protein 131–like, a regulator of thymocyte proliferation[5]. The A>G mutation causing a His320Arg change is predicted across *in silico* programs to be benign and very well tolerated. *INTS2* codes for a subunit of the Integrator protein complex which is impactful in gene regulation[6]. The C>T mutation causing a Arg421Cys change is predicted to be damaging and potentially deleterious. While a limitation of this study is that we have no concrete way to rule out either of these mutations as causative, neither of these genes have known functions related to the previously identified mechanism of harlequin.

Gene	Missense	PolyPhen	Function
BRCA1	Lys220Arg	0.978	E3 ubiquitin ligase, DNA repair
TMEM131L	His320Arg	0.000	T cell proliferation
INTS2	Arg421Cys	0.999	Integrator complex subunit

Table 1. The remaining three candidate variants, their PolyPhen scores, and functions.

The *BRCA1* mutation is of great interest since BRCA1 functions as an E3 ubiquitin ligase and serves a purpose in the ubiquitin-proteasome pathway. Little is known about the role of BRCA1 in ligation, and even less is known about the substrates with which it interacts[7]. We hypothesize that BRCA1 has an affinity for the SILV protein and is primarily responsible for the ubiquitination of aberrant SILV. If BRCA1 plays a role in this way, even a mild impairment of function could start the accumulation of mutant SILV and trigger melanocyte death. We do not believe this mutation would cause the death of other cells, since *SILV* is only expressed in melanocytes, and the excess mutant protein would be localized. While the cases in this study are limited to a singular family, we find it interesting to note that more than one family member, including the tri-colored sire, died at a young age from cancer. Due to the precedent set by the Great Dane mutation, we believe the *BRCA1* variant is the most likely genetic cause of the spontaneous harlequin coat patterning in this family of Australian shepherds. Future research efforts would include the synthesis of mutant SILV and the insertion of our variant into *BRCA1* to evaluate the ability of mutant E3 ligase to ubiquinate mutant SILV.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Leigh Anne Clark for her unwavering support and encouragement. She has been the best mentor I could have asked for and a much better one than I have ever deserved. I would like to thank Dr. Rooksana Noorai for her contributions to this project and generation of early data. Dr. Jacquelyn Evans also contributed to this work. I would also like to thank Drs. Alison Starr-Moss and Kate Tsai for their support of both me and the lab's work. Elizabeth Greif taught me many of the techniques I used for this project and provided input for much of my troubleshooting. Finally, I would like to thank the Clemson Honors College for the funding provided for this project.

REFERENCES

1. Clark LA, Tsai KL, Starr AN, Nowend KL, Murphy KE. A missense mutation in the 20S proteasome β2 subunit of Great Danes having harlequin coat patterning. *Genomics* 2011; **97**:244–8. https://doi.org/10.1016/j.ygeno.2011.01.003.
2. 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 USA* 2006; **103**:1376–81. https://doi.org/10.1073/pnas.0506940103.
3. Murphy SC, Evans JM, Tsai KL, Clark LA. Length variations within the Merle retrotransposon of canine PMEL: correlating genotype with phenotype. *Mob DNA* 2018; **9**:26. https://doi.org/10.1186/s13100-018-0131-6.
4. Tanaka K. The proteasome: overview of structure and functions. *Proc Jpn Acad Ser B Phys Biol Sci* 2009; **85**:12–36. https://doi.org/10.2183/pjab.85.12.
5. Maharzi N, Parietti V, Nelson E, Denti S, Robledo-Sarmiento M, Setterblad N, *et al.* Identification of TMEM131L as a Novel Regulator of Thymocyte Proliferation in Humans. *The Journal of Immunology* 2013; **190**:6187–97. https://doi.org/10.4049/jimmunol.1300400.
6. Wagner EJ, Tong L, Adelman K. Integrator is a global promoter-proximal termination complex. *Molecular Cell* 2023; **83**:416–27. https://doi.org/10.1016/j.molcel.2022.11.012.
7. Wu W, Koike A, Takeshita T, Ohta T. The ubiquitin E3 ligase activity of BRCA1 and its biological functions. *Cell Div* 2008; **3**:1. https://doi.org/10.1186/1747-1028-3-1.