5-2014

Discovery of Variants Underlying Simple and Complex Traits through Genome-Wide Association Study and Whole-Genome Resequencing

Rooksana Elizabeth Noorai

Clemson University

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

DISCOVERY OF VARIANTS UNDERLYING SIMPLE AND COMPLEX TRAITS
THROUGH GENOME-WIDE ASSOCIATION STUDY AND
WHOLE-GENOME RESEQUENCING

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Genetics

by
Rooksana Elizabeth Noorai
May 2014

Accepted by:
Dr. Leigh Anne Clark, Committee Chair
Dr. Susan C. Chapman
Dr. F. Alex Feltus
Dr. Keith E. Murphy
Dr. Michael G. Sehorn

ABSTRACT

If no reference genome exists, then the generation of a *de novo* genome assembly of an organism is necessary because having a reference genome expedites the discoveries for simple and complex traits. Reference genomes for the domestic dog (*Canis lupus familiaris*) and chicken (*Gallus gallus*) have resulted in the development of single nucleotide polymorphism (SNP) arrays for use in genome-wide association studies (GWAS). Next-generation sequencing technologies provide a rapid, increasingly affordable method for the generation of genome resequencing data.

The first objective of this work was to utilize existing genomic resources to investigate the genetic basis for traits of the dog and chicken. Episodic falling syndrome (EFS) is a recessive neurological disease of Cavalier King Charles spaniels. Using SNP profiles from only 12 individuals, EFS was mapped to chromosome 7; further experimentation led to the identification of the causative deletion. In a second example, SNP profiles from 197 German shepherd dogs were generated to identify loci underlying numerous diseases afflicting the breed, including recessive pituitary dwarfism, and three complex diseases: degenerative myelopathy, megaesophagus, and pancreatic acinar atrophy. Lastly, SNP profiles for 60 Araucana chickens were used to identify an association with the semi-dominant tailless rump (rumpless) phenotype on chromosome 2, as well as the recessive lethal ear-tufts phenotype on chromosome 15. Positional candidate genes were identified for both traits.

The second objective of this work was to identify loci associated with dermatomyositis (DM) and to develop resources to facilitate the identification of the causative mutation. DM is an inflammatory myopathy affecting humans and domestic dogs, primarily the collie and Shetland sheepdog breeds where painful lesions on the face and extremities are characteristics. The second objective was accomplished through 1) assembly of a population of DM-affected and healthy control collies, 2) completion of a GWAS using SNP profiles generated for this population, and 3) establishment of whole-genome resequencing data from 3 DM-affected collies and 2 healthy controls. Results revealed a strong association on chromosome 10. Annotation of the collie genome yielded novel SNPs, structural variants, and selective sweeps, and regions of reduced heterogeneity surrounding a gene(s) under strong positive selection.

DEDICATION

For Jessie and Ollie Noorai

ACKNOWLEDGMENTS

To be pasted.

TABLE OF CONTENTS

Table of Contents (Continued)

Table of Contents (Continued)

Table of Contents (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

List of Figures (Continued)

CHAPTER I

INTRODUCTION

OBJECTIVES OF THE DISSERTATION

My thesis is organized into two broad objectives. In the first I investigated several simple to complex traits (Chapters II-IV), and the second objective was to identify loci associated with dermatomyositis (DM) and to develop resources to facilitate the identification of the causative mutation (Chapters V and VI).

In the first objective I identified the genetic mutation that causes the autosomal recessive disease, Episodic falling syndrome, in the Cavalier King Charles spaniel (CKCS) breed (1). This was achieved by 1) using a genome-wide association study (GWAS) to identify the genetic region associated with the disease, 2) defining a haplotype present in affected CKCSs, 3) validate candidate genes to identify the causative mutation, a microdeletion in *BCAN*, present in affected and carrier CKCS, and 4) developing a genetic test to screen for carriers in the CKCS population (1) (Chapter II).

Next, I identified loci associated with four diseases (pituitary dwarfism, degenerative myelopathy, congenital megaesophagous (ME), and pancreatic acinar atrophy (PAA)) in the German shepherd dog (GSD) (2). The process for all four diseases was identical; 1) generating single nucleotide polymorphism (SNP) profiles for 197 dogs

and completing a GWAS for each disease and 2) determining haplotypes for each of the genetic regions identified. Candidate genes for pituitary dwarfism and degenerative myelopathy, *LHX3*and *SOD1*, were identified, respectively (2). ME and PAA proved to be complex diseases with multiple regions of association and no single causative mutation was identified (2) (Chapter III).

In the final part of the first objective, I identified the loci associated with the tailless and ear-tuft phenotypes in the Araucana chicken (3). This was achieved by conducting two GWASs using the same population of Araucana chickens and determining a haplotype for each of the genetic regions identified (3) (Chapter IV).

For the second objective, I studied dermatomyositis (DM), which is an autoimmune disease with a variable age of onset predominantly affecting collies (4). Although DM is known to be autosomal dominant with incomplete penetrance the causative mutation had not been identified (4). Using a GWAS with 27 DM affected and 20 controls collies, I identified a 10.5 Mb candidate region on chromosome 10 associated with DM. However, the coverage of the SNPs in the haplotype region was insufficient to discern a narrower region of interest. There are more than 130 identified genes through conservation of synteny with human chromosome 12 (5). However, only five homologous genes have been identified in the canine reference genome (Broad CanFam3.1) within the syntenic region (5) (Chapter V). The SNPs on the Illumina

2

CanineHD Infinium BeadChip are based on the most homogenous dog breeds, which currently excludes the collie (6).

With the long term aim of identifying the causative mutation of DM, I used next-generation sequencing (NGS) to resequence five collie dogs to map additional SNPs and structural variations unique to the collie, which is the final part of my thesis research. Five phenotypically diverse American collies (three DM and two control) from the original study were resequenced. Each collie was independently mapped to the canine reference genome to identify the polymorphisms and regions of reduced heterozygosity and structural variants unique to the collie breed. This was achieved by 1) generating 2x100 paired-end sequencing data; 2) aligning the remaining concordant paired sequence reads to the canine reference genome; and 3) identification of SNPS, structural variants common to all collies, and regions of reduced heterozygosity that could indicate the presence of selective sweeps in the collie breed.

I identified 9.7 million SNPs and more than 670 thousand structural variants (insertions, deletions, inversions, tandem repeats, and breakpoints) common to the collie genome in this cohort. Comparing the newly identified SNPs with the Illumina CanineHD Infinium BeadChip validated these data. Approximately 36,000 of these SNPs were predicted missense, splice site, frameshift, stop gained, and stop lost mutations. Finally, for validation of the structural variants within the resequenced data. I selected the largest structural variants in the following four classes: deletions, insertions, inversions and tandem duplications (Chapter VI).

THE RELATIONSHIP OF GENETIC VARIATION TO PHENOTYPE

Canine

The dog (*Canis lupus familiaris*) was the first animal to be domesticated by humans (7) and it is one of the most phenotypically diverse mammals (8-10). Through selection, humans have created isolated populations of each dog breed (11). In order for a dog to be a registered breed, both of its parents must be of the same registered breed. This is referred to as the pedigree barrier, which maintains the closure of each population (12,13). The existence of multiple dog breeds, a highly homozygous population unto themselves, makes the dog a valuable model in understanding the relationships between genotypes and phenotypes (7). The regions of homozygosity are up to 1×10^6 bp intrabreed and about 1×10^4 bp interbreed (6). SNPs are located throughout the genome. Because of the large regions homozygosity within a breed genomic changes exhibit significant statistical differences allowing for the identification of a large region(s) of association. The smaller regions of homozygosity between breeds are then used to narrow the candidate region. Thus, together, the regions within and between breeds are advantageous for identifying regions of association through the use of methods such as SNP array (BeadChip), GWAS, and fine mapping. (1,2,14,15-18).

The collie, the German Shepherd dog, and the Cavalier King Charles spaniel (Fig. 1) are examples of three breeds with genome variations leading to alternate phenotypic traits such as snout length and shape, fur length, body size and ear position. The collie originated from Scotland and Northern England, where it was selected for its sheep

4

herding abilities (19). The standard collie has either long or short fur, a distinctive long nose (dolichocephalic) and upright ears (15,19). The German Shepherd dog was originally a herding dog (20), but it is now prized for its aptitude for learning new skills, and protective nature (20). They have long fur, mid-length noses (mesocephalic) and tall pointed ears (15,20). The Cavalier King Charles spaniel however, is a type of English toy spaniel bred for companionship (21,22). These dogs have flat faces (brachycephalic), medium length fur, folded ears and a smaller physical size (22). These traits are examples of monogenic or polygenic genome variations.

Fig. 1. Diverse phenotypes of a collie, German Shepherd dog (GSD), and Cavalier King Charles spaniel (CKCS) (from left to right). The collie has a long nose, long or short fur, and upright ears. GSD have mid length noses, large pointed ears and long dark fur, whereas CKCS have short noses, floppy ears and are physically smaller (15,19-22).

Permissions for the images: 1. collie – Dr. Leigh Anne Clark, 2. GSD – Anne Skob, and 3. CKCS – the Rinz family.

SNP array (BeadChip) and GWAS have been the successfully applied as the standard approach to identify the regions of association for the above traits (23-25). Where fine mapping was employed the region of association was narrowed and in some cases the causative mutation identified. For example, snout length and shape is a polygenic trait with candidate regions on chromosomes 1 and 5 (23). The combinatorial effect and identity of causative gene(s) has yet to be determined (23). The genetic basis for long and short fur length is determined by a single SNP in exon 1 of *FGF5,* which was identified by a GWAS approach and fine mapping of candidate genes (24). Although a single *IGF1* allele acts as a major determinant of small dog body size (25), a number of other loci contribute modifier genes, including *GHR, HMGA2*, *SMAD2*, and *STC2,* which influence overall body size, indicating that control of body size is a polygenetic trait (26). Ear position, erect versus floppy ears, has been studied using a GWAS, however the causative mutation was not determined, rather a 100 kb region of strongest association on chromosome 10 near *MSRB3* was identified (23).

In addition, canines have more than 350 known hereditary diseases, many with similar clinical signs to humans (16), making the dog a promising model for the study of genetic diseases as different breeds are predisposed toward certain diseases (27-29). The increased homogeneity within a breed allows for the identification of a candidate association region. When several breeds suffer from the same disease, due to the reduced homozygosity between breeds, the region of interest can be significantly narrowed with the ultimate goal of identifying the causative mutation (7). For example, collies, Shetland

sheepdogs, and Labrador retrievers tend to get dermatologic diseases (4). Certain large breed dogs such as German shepherd dogs, Labrador retrievers, and Great Danes, have a tendency to develop hip dysplasia (30).

Identification of the genetic causes of disease has also used the standard SNP array, GWAS and fine mapping approach. However, the introduction of whole-genome resequencing using massive parallel sequencing (a.k.a. NGS) has recently proven to be a powerful method in identifying genomic variations that can then be associated with known disease phenotypes (31-33). Using the Korean Jindo dog, 339 SNPs in 222 genes were identified, of which the study confirmed 89 SNPs in genes previously associated with human disease (32). In future studies, SNPs and other genomic variations will be used to identify candidate mutations responsible for diseases with no known genetic basis in dogs and humans.

Avian

Studies on phenotypic variations in the chicken (*Gallus gallus*) using the SNP array and GWAS approach have been delayed due to the lack of a suitable SNP array and comprehensive comparative genomic tools (34). Only a few quantitative traits and morphological variants have been classically mapped to a single gene mutation (35). A major deterrent to design of a useful SNP array, i.e. at least 100k informative genetic markers, was the lack of sufficient suitable genetic markers to accommodate the diverse range of chicken breeds (>240) (34). Two proprietary low and medium density SNP

7

arrays (3,000 and 60,000 genetic markers) were recently developed based on 7 breeds (36-39).

In this study, I examined the genetic basis for the rumplessness and ear-tuft phenotypes in the Araucana chicken (3) using the Illumina 60 k SNP array. The interest in these phenotypes rests in the fact that they are two of the three required breed standard traits for chicken fanciers in the United States (3,40), the third trait being the pea comb (41). Moreover, the lack of the tail vertebrae and associated soft tissue structures and the craniofacial anomalies in the ear-tufted birds are potential models for human diseases. Using the SNP array and GWAS analysis, I identified the region of association for each phenotype. Both traits segregate independently in the population; therefore, we were able to use individuals from the same population for each GWAS (3). Rumplessness was associated with a 2.14 Mb region on chromosome 2; and the ear-tufts phenotype was associated with a 0.58 Mb region on chromosome 15 (3) (Chapter IV).

Since this study, a large collection of SNPs that segregate within >243 chicken populations, has been identified (34). In 2013, a high-density 600k SNP array was developed in conjunction with Affymetrix and it is the first commercially available SNP array for the chicken (34). Our study, among a handful of others, provided the proof of principle for the SNP array and GWAS approach in identifying the genetic basis of chicken traits and diseases (3,38,39,42,43).

TECHNIQUES

In this section I briefly discuss the main techniques used throughout this study. The first section covers genetic techniques: linkage disequilibrium mapping and genome-wide association studies. In the second section two genomic techniques used include microarray analysis and next-generation sequencing. Finally, I briefly discuss some of the limitations of these approaches.

Genetic – linkage disequilibrium mapping

Linkage disequilibrium (LD) mapping utilizes the nonrandom association between alleles at multiple loci, which may be on more than one chromosome, to find an association with the phenotype of interest (44). LD is defined as the frequency of a haplotype, i.e. co-inherited loci that do not exhibit the expected random recombination (44). This method is used to identify marker-trait associations (45). Unlike, linkage mapping, LD is measured in base pairs and is therefore a physical distance between defined genetic markers. The more homozygous a species the longer the region of homozygosity, up to 1×10^6 bp within dog breeds and about 1×10^4 bp between dog breeds (6). This means that a relatively low number of genetic markers are needed for a mapping study in dogs (46). Comparative genomics between the dog and humans (*Homo sapiens*) show that humans have considerably less LD compared to the dog (46,47), i.e. shorter regions of homozygosity. This means that humans are more heterogeneous than the dog, exhibiting increased genetic diversity, and require larger numbers of genetic markers for determining regions of association. In a further example, LD mapping and

9

GWAS (discussed below) was used to identify recombination occurring in mitochondrial DNA (mtDNA) in hominids (47). The combined LD mapping and GWAS data showed reduced LD within mtDNA in both humans and chimpanzees (47).

Genetic – genome-wide association study

A genome-wide association study (GWAS) uses a set of genetic markers, comparing two groups that vary by one phenotype, to identify an association between genetic markers and the phenotype (48). This method is important because it does not require previous knowledge of which genes or pathways the phenotype of interest is related to (49). Advantages of the association method are 1) that association can detect genes that have only a small effect on the phenotype, 2) that genetic markers closer to the causative mutation can be identified, and 3) that unrelated individuals can be compared (49), none of which is possible with the less informative classical linkage studies.

GWASs have been applied to humans, model organisms, and non-model organisms (48). The predominant genetic marker being used for these associations is the single nucleotide polymorphism (SNP) (49). Significant SNPs define a region of association identifying a locus. Validation of SNPs is done using direct sequencing to confirm that the SNP matches the association with the observed phenotype. Association regions indicate the approximate region in the genome where the causative mutation of the phenotype is located (50). GWAS in itself is unable to determine the nature of the causative mutation. The causative mutation within the region of association needs to be

identified by fine mapping.

Genomic – polymorphism detection – microarrays

The use of microarrays to detect polymorphisms utilizes tens of thousands to millions of oligonucleotide probes to hybridize target DNA to an array (51). The types of probes may be sequence-specific oligonucleotide such as in the Affymetrix GeneChip® assay (52). Two probes are designated as a set, and each set member will detect one allele of a SNP (Allele A or Allele B) present in a DNA sequence (52). Another method, used in Illumina Infinium BeadChip arrays, is to hybridize the fragments of sample DNA with oligonucleotides attached to microbeads in order to detect SNPs (53).

The use of DNA microarray as a means for SNP detection was initially used to identify the association of SNPs with cancer (53). SNP arrays generate profiles for affected and unaffected populations (54). By conducting statistical comparisons, the significant SNP(s) associated with the affected phenotype are identified (54). Since the actual locations of the SNPs in the genome are known, candidate chromosomal regions of interest can be analyzed to identify the genetic cause of the phenotype (54).

Genomic – polymorphism detection – next-generation sequencing

The introduction of massively parallel sequencing (a.k.a next-generation sequencing (NGS)) has enabled identification of enormous numbers of SNPs and copy number polymorphisms simultaneously (55). When compared to available reference

11

sequences that structural variations within an individual genome can be identified (56). Moreover, increasingly sophisticated algorithms have enabled the more accurate generation of de novo reference sequences (55). These techniques have the power to alter understanding of the mutational spectrum in model organisms and humans (55).

SNPs exist almost universally and have become the primary genetic marker for genetic and genomic analyses (57). Based on the reduced costs, in addition to improvements in next-generation sequencing (NGS) and the bioinformatics needed to process the data, many projects to identify SNPs in model and non-model organisms have been undertaken (57). There are 3 major NGS technologies, and, because of the dynamic nature of their maximum capabilities, the method chosen for a specific type of research project may vary (57). The 3 NGS technologies are: 454 or Roche sequencing, Illuminia sequencing, and Applied Biosystems or SOLiD sequencing (57). In a single day, each one of these platforms can generate upwards of 1 Gb of data (57).

In order to handle the amount and type of NGS data produced, many analysis software packages are available, including, free and commercial, compatible on various computer platforms, proprietary and flexible capabilities (57). Regardless of whichever NGS platform and analysis software are used, the intended objective is usually to conduct SNP discovery (57). SNP detection by NGS has been successful in humans, fruit flies, grains, vegetables, and other species (57).

The method by which SNPs are detected in NGS data depends on the presence or absence of a reference genome (57). With a reference genome, a NGS read that aligns with only a single mismatch to the reference genome will be called a SNP (57). Without a reference genome, NGS reads produce a *de novo* genome assembly, and then aligned reads are compared for a single mismatch (57). Both of these methods are theoretical since the actual process requires additional information: read depth, quality scores, and consensus base ratios (not discussed) (57).

Limitations of approaches

The above techniques have both strengths (as discussed above) and limitations. Linkage mapping has shortcomings in that crossover events are not random. The existence of one crossover can interfere with another one occurring in nearby proximity (58,59). Multiple crossover events can cause underestimation of the recombination frequency when such events occur between the two points being calculated (44). Another issue that affects linkage between two loci is the existence of recombination hotspots in eukaryotic chromosomes (60,61). Finally, in related individuals modes of inheritance (epistasis, pleiotropy) can adversely affect the accuracy of linkage. As an example, incomplete penetrance (where some individuals with the genotype do not display the phenotype) can lead to a false negative (62).

Linkage disequilibrium mapping has weaknesses related to population stratification. In population stratification, genetic markers may register a false positive

(an association) or false negative (no association) in heterogeneous backgrounds (63,64). Furthermore, as linkage disequilibrium measures the degree to which alleles at two loci are associated, in related populations linkage disequilibrium analysis requires multiple generations and large numbers of related individuals, to identify significant association with the phenotype due to the homogeneity within the population (65).

When SNPs are used as the genetic marker, they do not exist in complete independence (49). Although GWAS has been successful at identifying simple traits, only a limited number of complex traits have been identified due to the complexity of heritability (65). It is possible that rare variants, epistasis, epigenetics, and environmental factors are being overlooked (65).

As for SNP arrays, two major limitations are the presence of nonspecific hybridization and the extensive period required for hybridization (53). A further limitation of microarrays is that there are batch effects that complicate the ability to compare data, between experiments (66).

Although NGS is the most advanced technology, it too, has limitations. Variant discovery is reliant on there being sufficient depth of coverage to accurately determine variance (67). NGS platforms have biases in amplification from the PCR, resulting in bias in amplification of certain transcripts and the introduction of errors within the sequence (57).

14

ABBREVIATIONS

Table 1. Abbreviations used throughout the introduction.

Name	Abbreviation
Cavalier King Charles spaniel	CKCS
Dermatomyositis	DM
Genome-wide Association Study	GWAS
German Shepherd dog	GSD
Linkage disequilibrium	LD
Megaesophagus	ME
Mitochondrial DNA	mtDNA
Next-generation sequencing	NGS
Pancreatic acinar atrophy	PAA
Single nucleotide polymorphism	SNP

REFERENCES

1. Gill JL, Tsai KL, Krey C, Noorai RE, Vanbellinghen JF et al. (2011) A canine BCAN microdeletion associated with episodic falling syndrome. Neurobiol Dis. 45:130-136.

2. Tsai KL, Noorai RE, Starr-Moss AN, Quignon P, Rinz CJ et al. (2012) Genome-wide association studies for multiple diseases of German shepherd dogs. Mammalian Genome 23(1-2):203-211.

3. Noorai RE, Freese NH, Wright LM, Chapman SC, Clark LA (2012) Genome-wide association mapping and identification of candidate genes for the rumpless and ear-tufted traits of the Araucana chicken. PLoS ONE 7(7): e40974.

4. Hargis AM, Haupt KH, Hegreberg GA, Prieur DJ, Moore MP (1984) Familial canine dermatomyositis: Initial characterization of cutaneous and muscular lesions. The American Journal of Pathology 116: 234-244.

5. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH et al. (2002) The human genome browser at UCSC. Genome Res Jun 12(6):996-1006.

6. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog *Nature* Dec 8 438(7069):803-19.

7. Galibert F, André C (2007) The dog: A powerful model for studying genotype-phenotype relationships. Comparative Biochemistry and Physiology 3(1):67-77.

8. Hart BL (1995) Analysing breed and gender differences in behaviour. In: Serpell J, editor. The domestic dog: its evolution, behaviour and interactions with people. Cambridge: Cambridge University Press. pp. 65–77.

9. Coppinger R, Coppinger L (2001) Dogs: A New Understanding of Canine Origin, Behavior and Evolution. Chicago: University of Chicago Press. 352 p.

10. Young A, Bannasch D (2006) Morphological variation in the dog. In: Ostrander EA, Giger U, Lindblad-Toh K, editors. The genome of the domestic dog. New York; Cold Spring Harbor Press. pp. 47–65.

11. Megens HJ, Groenen MAM (2012) Domesticated species form a treasure-trove for molecular characterization of Mendelian traits by exploiting the specific genetic structure of these species in across-breed genome-wide association studies. Heredity 109:1–3.

12. Mosher DS, Spady TC, Ostrander EA (2009) Dog In: Crockett NE, Chittaranjan K editors. Genome Mapping and Genomics in Animals, Volume 3 Heidelberg: Springer 231-256 pp.

13. Patterson DF (2000) Companion animal medicine in the age of medical genetics. Journal of Veterinary Internal Medicine 14**:**1-9.

14. Awano T, Johnson GS, Wade CM, Katz ML, Johnson GC et al. (2009) Genome-wide association analysis reveals a SOD1 mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis. Proc Natl Acad Sci USA 106: 2794–2799.

15. Schoenebeck JJ, Hutchinson SA, Byers A, Beale HC, Carrington B, et al (2012) Variation of *BMP3* contributes to dog breed skull diversity. PLoS Genet 8(8):e1002849.

16. Shearin AL, Ostrander EA (2010) Canine Morphology: Hunting for Genes and tracking mutations. PLoS Biology 8(3):e1000310.

17

17. Pemberton TJ, Choi S, Mayer JA, Li FY, Gokey N, et al. (2014) A mutation in the canine gene encoding folliculin-interacting protein 2 (FNIP2) associated with a unique disruption in spinal cord myelination. Glia 62(1):39-51.

18. Olsson M, Tintle L, Kierczak M, Perloski M, Tonomura N, et al. (2013) Thorough Investigation of a Canine Autoinflammatory Disease (AID) Confirms One Main Risk Locus and Suggests a Modifier Locus for Amyloidosis. PLoS ONE 8(10):e75242.

19. American Kennel Club (2013) Collie: History. Available: http://www.akc.org/breeds/collie/history.cfm. Accessed 29 November 2013.

20. American Kennel Club (2013) German Shepherd Dog: History. Available: http://www.akc.org/breeds/german_shepherd_dog/history.cfm. Accessed 29 November 2013.

21. Cavalier King Charles Spaniel Club – USA (2013) Cavalier: A natural healer. Available: http://www.ckcsc.org/ckcsc/ckcsc_inc.nsf/Founded-1954/therapy.html. Accessed 29 November 2013.

22. Cavalier King Charles Spaniel Club – USA (2013) History of the cavalier king Charles spaniel. Available: http://www.ckcsc.org/ckcsc/ckcsc_inc.nsf/Founded-1954/breedhist.html. Accessed 29 November 2013.

23. Byoko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, et al. (2010) A simple genetic architecture underlies morphological variation in dogs. PLoS Biology 8(8): e1000451.

24. Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, et al. (2009) Coat variation in the domestic dog is governed by variants in three genes. Science 326(5949):150-153.

25. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K et al. (2007) A single IGF1 allele is a major determinant of small size in dogs. Science 316:112–115.

26. Rimbault M, Beale HC, Schoenebeck JJ, Hoopes BC, Allen JJ, et al. (2013) Derived variants at six genes explain nearly half of size reduction in dog breeds. Genome Research 23(12):19985-1995.

27. Giger U, Sargan DR, McNiel EA (2006) Breed-specific hereditary diseases and genetic screening. In: Ostrander EA, Giger U, Lindblad-Toh K, editors. The dog and its genome. New York: Cold Spring Harbor Laboratoary Press. pp. 249-289.

28. Sargan DR (2004) IDID: inherited diseases in dogs: web-based information for canine inherited disease genetics. Mamm Genome 15:503-6.

29. Patterson DF (2000) Companion animal medicine in the age of medical genetics. J Vet Intern Med 14:1-9.

30. Mateescu RG, Dykes NL, Todhunter RJ, Acland GM, Burton-Wurster NA, et al. (2006) QTL mapping using crossbreed pedigrees: Strategies for canine hip dysokasua. In: Ostrander EA, Giger U, Lindblad-Toh K, editors. The dog and its genome. New York: Cold Spring Harbor Laboratoary Press. pp. 407-438.

31. Stothard P, Choi JW, Basu U, Sumner-Thomson JM, Meng Y et al. (2011) Whole genome resequencing of Black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics. 12:559-572.

32. Kim RN, Kim DS, Choi SH, Yoon BH, Kang A et al. (2012) Genome Analysis of the Domestic Dog (Korean Jindo) by Massively Parallel Sequencing**.** DNA Research. 19(3):275-288.

33. Fan WL, Ng CS, Chen CF, Lu MJ, Chen YH et al. (2012) Genome-wide patterns of genetic variation in two domestic chickens. Genome Biology and Evolution. 5(7):1376-1392.

34. Kranis A, Gheyas AA, Boschiero C, Turner F, Le Y, et al. (2013) Development of a high density 600K SNP genotyping array for chicken. BMC Genomics 14:59-71.

35. Tixier-Boichard M (2002) From phenotype to genotype: Major genes in the chickens. World's Poultry Science Journal 58:35-45.

36. International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432(7018):695-716.

37. Muir WM, Wong GK, Zhang Y, Wang J, Gorenen MAM, et al. (2008) Review of the initial validation and characterization of a chicken 3K SNP array. World's Poultry Science Journal 64:219-226.

38. Groenen MAM, Megens HJ, Zare Y, Warren WC, Hillier LW, et al. (2011) The development and characterization of a 60K SNP chip for chicken. BMC Genomics 12: 274.

39. Robb EA, Gitter CL, Cheng HH, Delany ME (2011) Chromosomal mapping and candidate gene discovery of chicken developmental mutants and genome-wide variation analysis of MHC congenics. J Hered 102(2):141-156.

40. The Araucana Club of America (2011 Jan 18) The araucana club of America. Available: http://www.araucana.net/. Accessed 29 November 2013.

41. Boije H, Harun-or-Rashid M, Lee YJ, Imsland D, Bruneau N, et al. (2012) Sonic Hedgehog-signalling patterns the developing chicken comb as revealed by exploration of the pea-comb mutation. PLoS ONE 7(12):e50890.

42. Liu R, Sun Y, Zhao G, Wang F, Wu D, et al. (2013) Genome-wide association study identifies Loci and candidate genes for body composition and meat quality traits in Beijing-You chickens. PLoS ONE 8(4):e61172.

43. Pettersson ME, Johansson AM, Siegel PB, Carlborg O (2013) Dynamics of Adaptive Alleles in Divergently Selected Body Weight Lines of Chickens. G3 (Bethesda) 3(12):2305–2312.

44. Griffiths AJF, Wessler SR, Lewontin RC, Carroll SB (2008) Introduction to genetic analysis ninth edition. New York: W. H. Freeman and Company 838 p.

45. Integrated breeding platform (2013) Linkage disequilibrium mapping – LD mapping. Available: https://www.integratedbreeding.net/linkage-disequilibrium-mapping-ld-mapping. Accessed 27 November 2013.

46. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, et al. (2004) Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. Genome Res 14:2388-2396.

47. Awadalla P, Eyre-Walker A, Smith JM (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science 286:2524-2525.

48. Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29-37.

49. Dick D (2008) Introduction to association. In: Neale BM, Ferrerira MAR, Medland SE, Posthuma D, editors. Statistical genetics: Gene mapping through linkage and association. New York: Taylor & Francis Group. pp. 311-321.

50. National Human Genome Research Institute: National Institutes of Health (2013, Jul 11) Genome-wide association studies. Available: http://www.genome.gov/20019523. Accessed 27 November 2013.

51. NCBI (2013) Microarrays. Available: http://www.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechMicroarray.shtml. Accessed 28 November 2013.

52. NCBI (2013) Sequence-specific oligonucleotide (SSO) probes. Available: http://www.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechSSO.shtml. Accessed 28 November 2013.

53. Bromberg A, Jensen EC, Kim J, Kyung Y, Mathies RA (2012) Microfabricated linear hydrogel microarray for single-nucleotide polymorphism detection. Analytical Chemistry 84(2):963-970.

54. Chial H (2008) Rare genetic disorders: Learning about genetic disease through gene mapping, SNPs, and microarray data. Nature Education 1(1):192.

55. Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends in Genetics 24(3):133-141.

56. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 25(21): 2865-2871.

57. Kumar S, Banks TW, Cloutier S (2013) SNP discovery through next-generation sequencing and its applications. International Journal of Plant Genomics 2012:831460-831474.

58. Fagerness J, Nyholt DR (2008) Basics of DNA and genotyping. In: Neale BM, Ferrerira MAR, Medland SE, Posthuma D, editors. Statistical genetics: Gene mapping through linkage and association. New York: Taylor & Francis Group. pp. 5-16.

59. Robine N, Uematsu N, Amiot F, Gidrol X, Barillot E, et al. (2007) Genome-Wide Redistribution of Meiotic Double-Strand Breaks in *Saccharomyces cerevisiae. Molecular and Cellular Biology* 27(5):1868-1880.

60. Nelson DL, Cox MM (2008) Lehninger principles of biochemistry fifth edition. New York: W.H. Freeman and Company. 1158 p.

61. Paigen K, Petko P (2010) Mammalian recombination hot spots: properties, control and evolution. Nature Reviews Genetics 11:221-233.

62. Nyholt DR (2008) Principles of linkage analysis. In: Neale BM, Ferrerira MAR, Medland SE, Posthuma D, editors. Statistical genetics: Gene mapping through linkage and association. New York: Taylor & Francis Group. pp. 113-134.

63. Wray NR, Visscher PM (2008) Population genetics and its relevance to gene mapping. In: Neale BM, Ferrerira MAR, Medland SE, Posthuma D, editors. Statistical genetics: Gene mapping through linkage and association. New York: Taylor & Francis Group. pp. 87-112.

64. Hoffman GE (2013) Correcting for population structure and kinship using the linear mixed model: theory and extensions. PLoS ONE 8(10):e75707.

65. Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. Lancet 361:598-604.

66. Gibson G (2010) Hints of hidden heritability in GWAS. Nature Genetics. 42(7):558-560.

67. Reese SE, Archer KJ, Therneau TM, Atkinson EJ, Vachon CM, et al. (2013) A new statistic for identifying batch effects in high-throughput genomic data that uses guided principal component analysis. Bioinformatics. 29(22):2877-2883.

68. Xu F, Wang W, Wang P, Li MJ, Sham PC, et al. (2012) A fast and accurate SNP detection algorithm for next-generation sequencing data. Nature Communications 3:1258-1266.

CHAPTER II

A CANINE *BCAN* MICRODELETION ASSOCIATED WITH EPISODIC FALLING SYNDROME

Jennifer L. Gill[a], Kate L. Tsai[b], Christa Krey[c], Rooksana E. Noorai[b], Jean-François Vanbellinghen[d], Laurent S. Garosi[e], G. Diane Shelton[f], Leigh Anne Clark[b], Robert J. Harvey[a]

[a]Department of Pharmacology, The School of Pharmacy, 29-39 Brunswich Square, London WC1N 1AX, UK

[b]Department of Genetics and Biochemistry, College of Agriculture, Forestry, and Life Sciences, 100 Jordan Hall, Clemson University, Clemson, South Carolina 29634-0318, USA

[c]55 Bruce Road, Levin 5510, New Zealand

[d]Biologie Moléculaire, Institut de Pathologie et de Génétique ASBL, 25 Avenue Georges Lemaître, B-6041 Gosselies, Belgium

[e]Davies Veterinary Specialists, Manor Farm Business Park, Higham Gobion, Hertfordshire, UK

[f]Department of Pathology, University of California, San Diego, La Jolla, CA 92093-0709, USA

Permission to reprint was granted, please see appendix D.

ABSTRACT

Episodic falling syndrome (EFS) is a canine paroxysmal hypertonicity disorder found in Cavalier King Charles spaniels. Episodes are triggered by exercise, stress or excitement and characterized by progressive hypertonicity throughout the thoracic and pelvic limbs, resulting in a characteristic 'deer-stalking' position and/or collapse. We used a genome-wide association strategy to map the EFS locus to a 3.48 Mb critical interval on canine chromosome 7. By prioritizing candidate genes on the basis of biological plausibility, we found that a 15.7 kb deletion in *BCAN*, encoding the brain-specific extracellular matrix proteoglycan brevican, is associated with EFS. This represents a compelling causal mutation for EFS, since brevican has an essential role in the formation of perineuronal nets governing synapse stability and nerve conduction velocity. Mapping of the deletion breakpoint enabled the development of Multiplex PCR and Multiplex Ligation- dependent Probe Amplication (MLPA) genotyping tests that can accurately distinguish normal, carrier and affected animals. Wider testing of a larger population of CKCS dogs without a history of EFS from the USA revealed that carriers are extremely common (12.9%). The development of molecular genetic tests for the EFS microdeletion will allow the implementation of directed breeding programs aimed at minimizing the number of animals with EFS and enable confirmatory diagnosis and pharmacotherapy of affected dogs.

INTRODUCTION

Episodic falling syndrome (EFS) is a well-recognized paroxysmal disorder found in Cavalier King Charles spaniels (CKCS). Episodes begin between fourteen weeks and four years of age and are triggered by exercise, stress, apprehension or excitement (1). Episodes are of variable frequency and severity but are characterized by progressive hypertonicity involving thoracic and pelvic limbs (Fig. 1a) until the dogs are ultimately immobilized in a characteristic 'deer-stalking' or 'praying' position (Fig. 1b). Stiffening of all four limbs during exercise can cause falling (Fig. 1c), although there is no loss of consciousness or cyanosis. Other clinical signs may include facial muscle stiffness, stumbling, a 'bunny-hopping' gait, arching of the back or vocalization. Curiously, between episodes, dogs appear to be completely normal neurologically. Spontaneous activity was not observed in muscle electrodiagnostic testing, ruling out myotonia congenita (2,3). Muscle biopsies are typically normal at the light microscopic level, excluding many congenital myopathies.
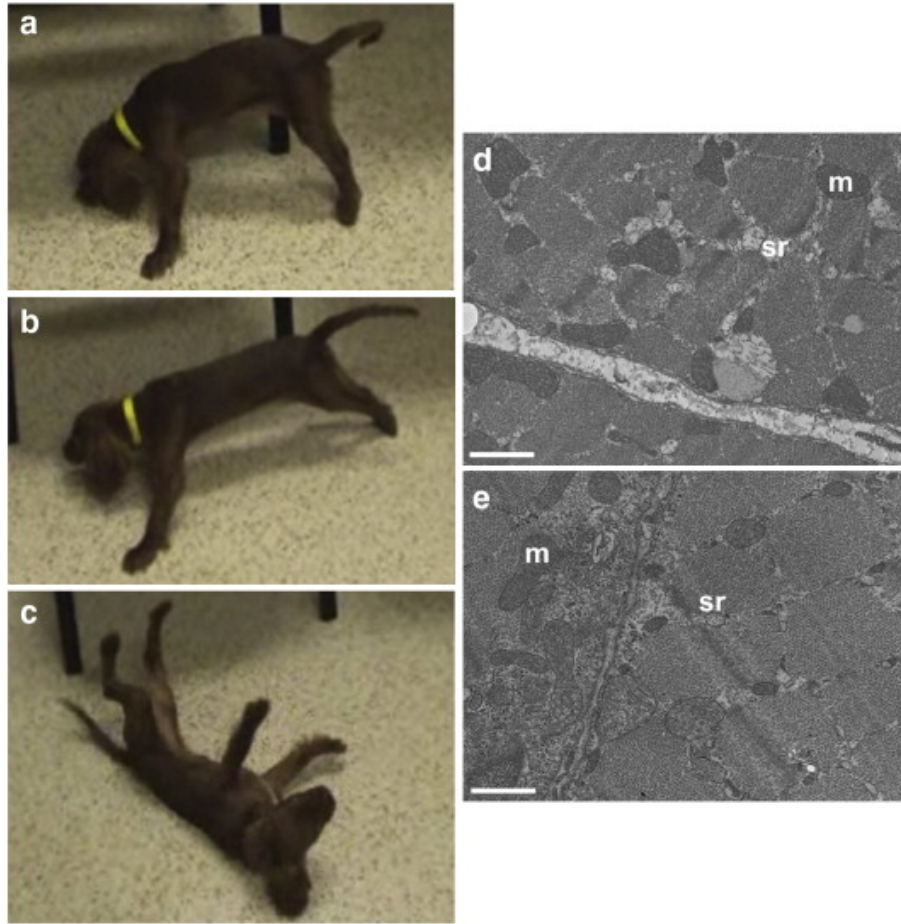
Fig. 1. Clinical signs of episodic falling syndrome and muscle pathology. A 5-month-old female Cavalier King Charles spaniel presented with typical episodes of excitement or exercise-induced muscle stiffness (a, hypertonicity) that would involve all four limbs and progress to an usual 'deer-stalking' or 'praying' posture (b), eventually resulting in falling (c). While EFS muscle was normal histologically by light microscopy, electron microscopy (d) revealed that the sarcoplasmic reticulum (sr) appeared dilated and contained finely granular material compared to control muscle (e). Mitochondria (m) and myofibrils were normal in appearance in both tissues. Scale bars = 0.31 μm.

However, EFS has been linked to ultrastructural defects in skeletal muscle including dilatation and proliferation of the sarcoplasmic reticulum, mitochondrial swelling and degeneration (2,3). EFS has also been compared (4) with startle disease/hyperekplexia, typically characterized by noise- or touch-evoked neonatal hypertonicity due to defects in inhibitory glycine receptor (*GLRA1*, *GLRB*; 5,6) or

glycine transporter GlyT2 (*SLC6A5*) genes (7,8). However, a microdeletion in the GlyT2 gene in Irish Wolfhounds results in severe neonatal muscle stiffness and tremor in response to handling (9), which is inconsistent with the observed clinical signs of EFS. Comparisons with startle disease may have been made because affected dogs often respond well to the benzodiazepine clonazepam (10), an effective anticonvulsant, anxiolytic and muscle relaxant that is the most effective known treatment for human hyperekplexia (11). However, the carbonic anhydrase inhibitor acetazolamide, used to treat certain types of human episodic ataxia (12) and hyperkalemic periodic paralysis (13), also appears to have therapeutic value in the treatment of EFS (http://www.cavalierhealth.org/episodic_falling).

Since a ten-year breeder-led investigation into the inheritance of EFS suggested an autosomal recessive mode of inheritance (http://cavalierepisodicfalling.com/), we used a genome-wide association strategy (14) to map the EFS locus to a defined region of canine chromosome 7. Candidate gene analysis enabled us to identify a microdeletion affecting the brevican gene (*BCAN*), confirm the deletion breakpoint and develop rapid genotyping tests for EFS.

MATERIALS AND METHODS

Light and electron microscopy:

For light microscopy, unfixed biopsies from the biceps femoris, vastus lateralis and triceps brachii muscles were collected from five affected CKCS dogs under general

29

anesthesia and frozen in isopentane pre-cooled in liquid nitrogen. Cryosections were cut (8 μm) and the following histochemical stains and reactions performed: hematoxylin and eosin, modified Gomori trichrome, periodic acid Schiff, phosphorylase, esterase, ATPase reactions at pH of 9.8 and 4.3, nicotinamide adenine dinucleotide-tetrazolium reductase, succinic dehydrogenase, acid phosphatase, alkaline phosphatase and oil red O. For electron microscopy, glutaraldehyde-fixed muscle specimens were post-fixed in osmium tetroxide, and dehydrated in serial alcohol solutions and propylene oxide prior to embedding in araldite resin. Thick sections (1 μm) were stained with toludine blue for light microscopy and ultrathin sections (60-90 μm) were stained with uranyl acetate and lead citrate for electron microscopy.

Study cohort and DNA preparation:

Our study cohort comprised: EFS affected - 10 animals (6 from the USA, 2 from New Zealand and 2 from the UK); Obligate EFS carriers - 8 animals (2 from the USA, 6 from New Zealand); Animals related to carriers or affected dogs - 21 animals (7 from the USA, 14 from New Zealand); Controls - CKCS with no EFS history - 14 animals (all from the USA). Genomic DNA was isolated from whole blood or buccal cells using the Gentra Puregene Blood Kit (QIAGEN, Valencia, USA). Additional DNA samples from 155 CKCS with no clinical history of EFS and other pure bred-dogs were available from unrelated studies and other sources (e.g.

Cornell DNA bank: http://www.vet.cornell.edu/research/dnabank/).

Genome-wide association mapping:

Thirteen CKCS genomic DNA samples isolated from blood (five cases, one obligate carrier and seven controls from the USA) were genotyped for 127,000 SNPs on the Affymetrix Canine SNP Array version 2 (http://www.broadinstitute.org/mammals/dog/caninearrayfaq.html). The two main drivers for sample selection were: i) lack of relatedness - i.e., that the animals used for case-control analysis should not share a common ancestor within at least three generations and ii) the quality and quantity of genomic DNA available. Arrays were processed at the Clemson University Genomics Institute (http://www.genome.clemson.edu/) using the GeneChip human mapping 250K Sty assay kit (Affymetrix, Santa Clara, USA). The GeneChip human mapping 500K assay protocol was followed, but with a hybridization volume of 125 μl (14). Raw CEL files were genotyped using Affymetrix Power Tools software. SNPs having >10% missing data and ≥60% heterozygosity were removed. Data for 58,873 SNPs were formatted for PLINK (15) and case/control analyses with 100,000 permutations were performed for five cases and seven controls (the obligate carrier was excluded from analysis).

PCR and DNA sequencing:

PCR primers were designed to amplify exons and flanking splice donor, acceptor and branch-point sites, from gene structures derived *in silico* using the UCSC Genome Browser. For exon-specific primers for *BCAN* and *HAPLN2* exon amplification see Table S1. PCR was performed using 50 ng genomic DNA as template and AccuPrime *Pfx*

31

SuperMix supplemented with betaine for 40 cycles of 94°C for 1 min, 60°C for 1 min, 68°C for 1 min. PCR products were gel purified using a QiaQuick gel extraction kit (QIAGEN, Crawley, UK) for TOPO cloning (pCR4Blunt-TOPO; Invitrogen, Renfrew, UK). Sanger DNA sequencing was performed by DNA Sequencing & Services (MRCPPU, College of Life Sciences, University of Dundee, Scotland) using Applied Biosystems Big-Dye version 3.1 chemistry on an ABI 3730 automated capillary DNA sequencer. DNA sequences were analyzed using Sequencher 4.10 (Gene Codes Corporation, Ann Arbor, USA). For multiplex PCRs, the diagnostic primer set is: EFS1 5'-aaggtcttacacctgcaatgaatag-3', EFS2 5'-agcaaatgtaaagtcctgtgaccat-3' and EFS3 5'-agttcacattgtgctctctctactg-3'.

Multiplex ligation-dependent probe amplification (MLPA) analysis:

Five MLPA probe sets were designed corresponding to the promoter region (PR) and exons 1 (5' UTR), 2, 3 and 4 of the canine brevican gene (*BCAN*; NC_006589 on chromosome 7; Table S2). Criteria for MLPA probe design were as previously described (Schouten *et al*. 2002). A control probe pair was designed to recognize an unrelated gene (CFTR: NC_006596 on chromosome 14). Probes generated amplification products ranging in size from 88 to 115 bp and had annealing temperatures higher than 70°C as recommended in RAW Probe software (MRC-Holland, Amsterdam, The Netherlands) using standard MLPA conditions (Schouten *et al*. 2002). PCR products were analyzed on an ABI 3130XL capillary electrophoresis apparatus (Applied Biosystems, Lennik, Belgium). Normalization of *BCAN-* specific probe signals was performed by dividing the

32

values obtained by the combined signal of the control probes.

RESULTS

Light and electron microscopy:

Unfixed cryosections of muscle biopsies were histologically normal at the light microscopic level with no abnormalities detected following any of the histochemical stains and enzyme reactions employed. Contrary to previous reports (2,3) electron microscopy revealed normal myofibrillar and mitochondrial morphology, although swelling of the sarcoplasmic reticulum was confirmed (Fig. 1d,e).

Genome-wide association mapping and candidate gene resequencing:

A total of 17 single nucleotide polymorphisms (SNPs) were associated with EFS (Praw values ≤ 0.0001) (Table S3). The most significant result was for SNP 43389066 (Praw = 5.10×10^{-7}, Pgenome = 2.68×10^{-3}) (Fig. 2a). All significant SNPs were located within a 7.2 Mb region on canine chromosome 7. A critical interval of 3.48 Mb (from 7.42838021 to 7.46320904) was delimited by recombinant chromosomes identified in one EFS dog and an obligate carrier (Fig. 2b). In order to identify the mutation associated with EFS, we prioritized several genes for resequencing based on biological plausibility. These encoded ligand-gated or voltage-gated ion channels (*CHRNB2*, *HCN3*, *KCNN3*), mitochondrial (*MRPL24*, *MTSO1*, *MTX1, SLC25A44*), muscle (*MEF2D*, *TPM3*) or brain-expressed proteins (*ARHGEF11*, *BCAN*, *GBA*, *HAPLN2*, *NES*, *RIT1*, *SYT11, UBQLN4*). Curiously, amplification of *BCAN* exons 1, 2 and 3 consistently failed with multiple

33

primer sets when using genomic DNA from affected animals, while DNAs from carriers and unaffected dogs amplified reliably (Fig. 3a). Because no preferential amplification was observed for the adjacent gene, *HAPLN2* - encoding hyaluronan and proteoglycan link protein 2/Bral1 (16-18) - we suspected that a microdeletion affecting *BCAN* regulatory sequences and exons 1-3 was associated with EFS.
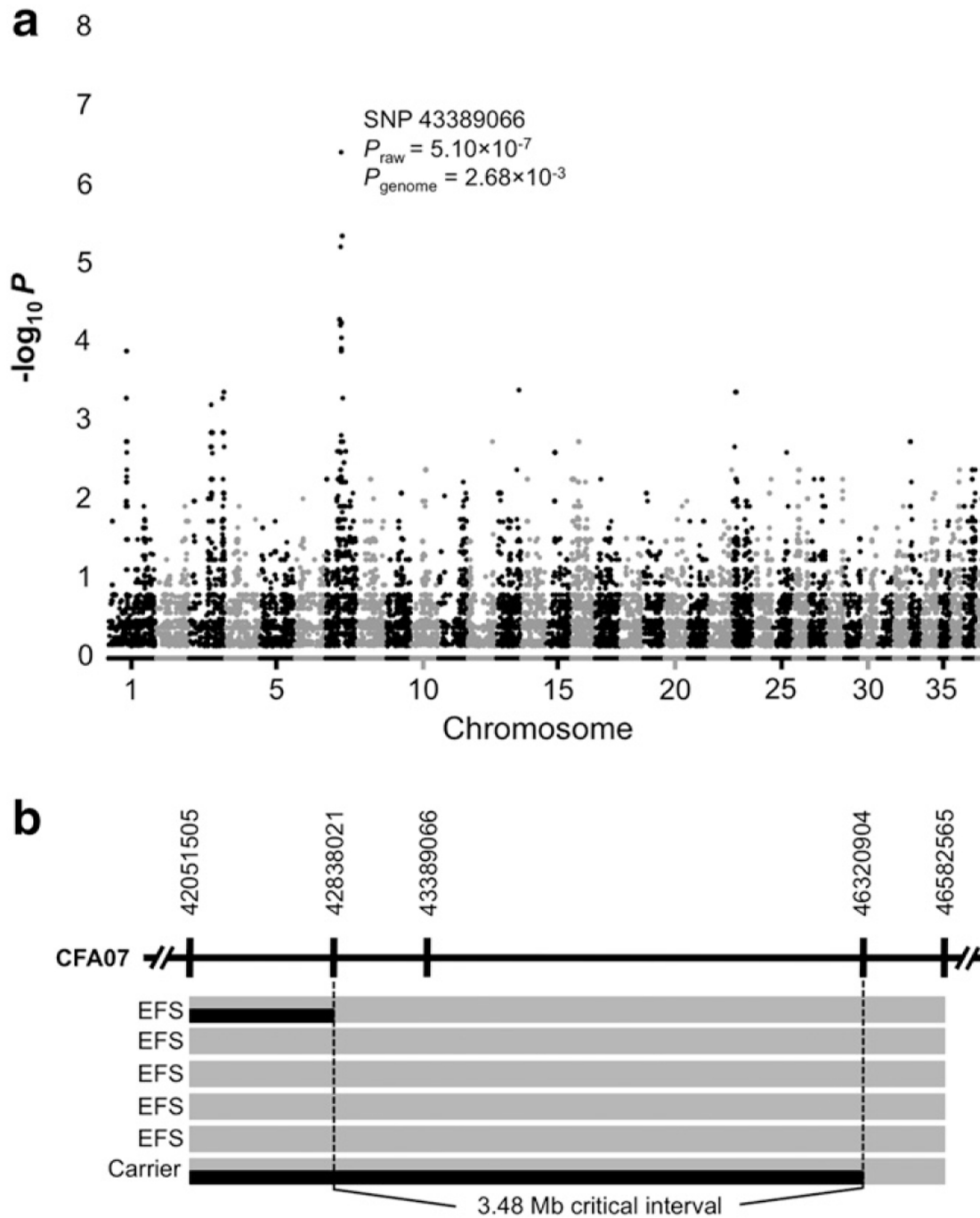
Fig. 2. Mapping the episodic falling syndrome locus. (a) $-\log_{10}$ of P_{raw} values (Y axis) for genome-wide association using five EFS and seven control dogs are plotted for each chromosome (X axis). A single major signal was detected on chromosome 7. (b) A 3.48 Mb critical interval encompassing SNP 43389066 is defined by recombination events in an affected CKCS and an obligate carrier.

Deletion breakpoint identification and development of diagnostic tests:

Further primer walking experiments enabled us to clone and sequence a DNA fragment containing the deletion breakpoint and develop a multiplex PCR assay that distinguishes between affected, carrier and normal dogs (Fig. 3b). Sequence analysis of the breakpoint amplicon revealed a 15.7 kb microdeletion starting 1.56 kb downstream of *HAPLN2,* encompassing *BCAN* promoter elements and exons 1 (5' untranslated), 2 and 3, finishing 85 bp downstream of *BCAN* exon 3 (Fig. 3c). The microdeletion amplicon also contained a 6 bp inserted sequence (GGCCTT; Fig. 3d) typical of deletions resulting from non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ). Several regions of microhomology (1-7 bp) were identified in a 30 bp region encompassing the breakpoints (Fig. 3e). Interestingly, 5 of 6 bp of the reverse complemented inserted sequence aligns to the largest region of microhomology. We also noted an abundance of short interspersed element (SINE) insertions at the 5' end of the deleted sequence, which could cause the formation of secondary structures that facilitate chromosomal rearrangement (19). We also detected the presence of the *BCAN* microdeletion (Fig. 4) using MLPA (20) and canine-specific probe sequences (Table S3).
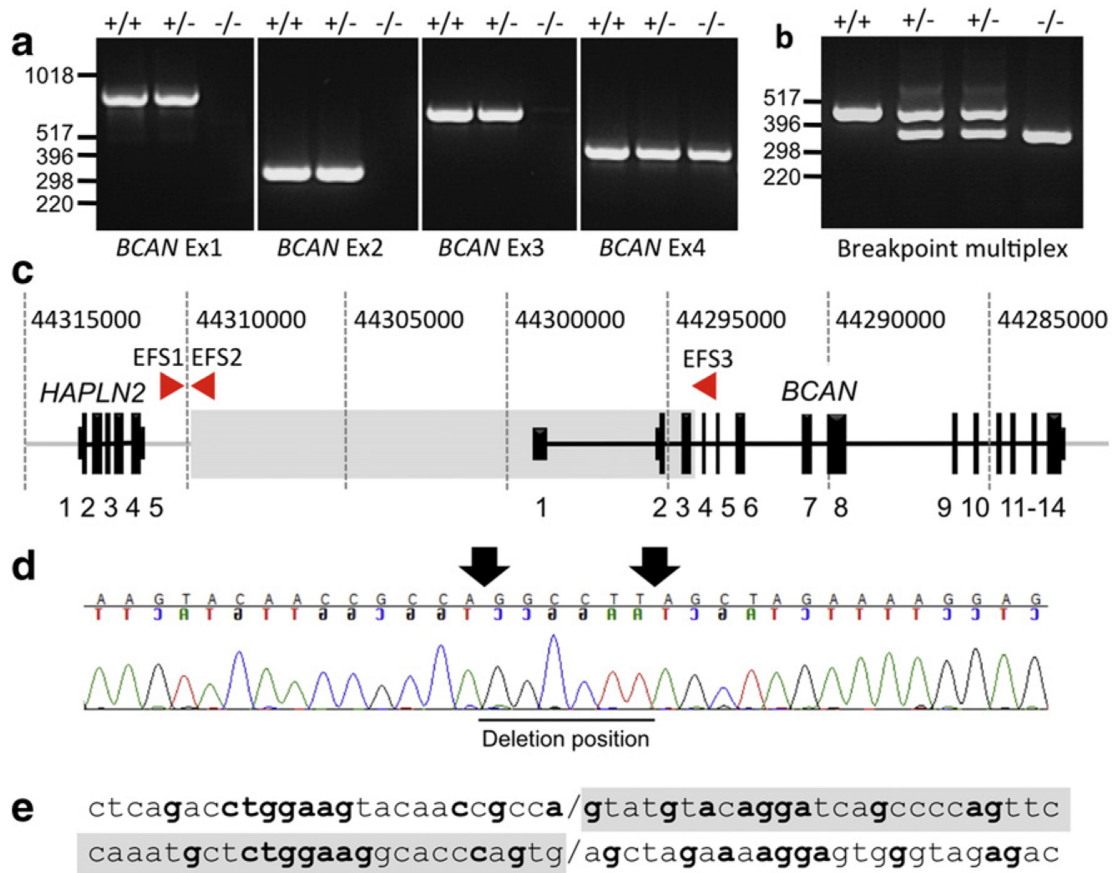
Fig. 3. (a) PCR panels for BCAN exons 1–4 showing that amplicons for the first three exons of the brevican gene can be generated from genomic DNA from normal (+/+) or obligate carrier (+/-) samples, but cannot be amplified from an equivalent EFS sample (-/-). By contrast, BCAN exon 4 can be amplified from all genotypes. (b) Multiplex PCRs with primers flanking the 15.7 kb BCAN microdeletion allowed simultaneous detection of the wild-type BCAN allele (primers EFS1 + 2, 393 bp) in normal (+/+) or EFS carrier (+/-) animals, while the EFS allele (primers EFS1 + 3, 273 bp) is detected in both EFS carrier and affected (-/-) dogs. Note that the two carriers shown have both wild-type and EFS amplicons, as expected for a heterozygous genotype. (c) Schematic diagram showing the genomic organization of the HAPLN2 and BCAN, the position of the deletion (grey shading) and EFS1-3 primers. (d) Sequence spanning the BCAN deletion breakpoint, showing an additional non-homologous inserted sequence indicated by arrows. (e) Alignment of the DNA sequence immediately flanking the deletion breakpoint indicating local microhomology.

```
cagaagggaaaagtaagaagagctctgaaaggtcttacacctgcaatgaatagctgcagcgcagagagac
acctgtctctcccctcgacagctgtttctctagaacgaatcccatggccttgcttaaccaaaaggggca
gaaaagagcaacctactcagacctggaagtacaaccgccagtatgtacaggatcagccccagttccacca
caggcatcagggcgggcagactggtgggcacccactcctccccggctcttctgggtttttctgagttacag
tctgtacggagtcagagccccagaaagccaagacgctggacccagaagcttccctgaaggaagatggctt
.............................10430bp.............................
gcactcgctcccgccctccagccctcccccctactccaagtccccggcttctctccagaaccatgtcct
acttcgcacgtccttgctctctctgctgcctctaggaaccgacgcagaggaggcagcggtagcgtgacct
tcccgagcccctgcgtccctggacaggggggcggggcggccccgggggaggagcggggcggggacgggcg
ggggagtggagaaaggggttttgtgcgagcccggcggcccggcgccctcttccgaacgtcctgcggcgcc
ggccctccctgcgcccgcgcagtcctcgccgcagtcggccgcagccgcaggacggagcgtggaccgggag
gagagcccggaggaggctgcaaacttggcggtgcagcaccctcgcagccccggccgcctgcaggtaggtg
gagcacgagcgcagccggaccaccgggacccgagaggcagggagcccggggcgccgtcggccgccggtgg
.............................3430bp.............................
agtccacatggatgtggtcaggagccggggggaagtccatcctaaagtctaaccccagctttcctcctcaa
gtccctccaccagcctgagcATGGCCCCCCTGTTCCTGCCCCTGCTGGTAGCCTTGGCCCTGGCCCCGGGG
CCCTGTGGCCTTAGCTGATGCCCTGGAAGGGGACAGCTCAGgtaagcaaccgcactcagcatcactgtct
ctgtctcctgcatctcctctatggaagagggtggccttggaggacccgggggagagataggatagagggca
.............................350bp.............................
tggggccctgggagcagtggtggctgggatcctggggcgggcttggagaccaggtgggaccctgtggatc
aggtgccaagttctagccaccggcctgcccgcccacccacagAGGACCGGGCCTTCCACGTGCGCATCGC
GGGTGACGCACCACTGCAGGGCGTGCTGGGCGGCGCCCTCACCATCCCTTGCCACGTTCACTACCTACGG
CCGCTGCCGGGCCACCGGGCCGTGCTGGGCTCCCCGCGGGTCAAGTGGACCTTCCTGTCCGGGGGCCGTG
AGGCCGAAGTATTGGTGGCTCGGGGGGCTGCGCGTCAAGGTGAGCGAGGCCTACCGTTTCCGTGTGGCACT
GCCTGCCTACCCGGCATCACTCACCGACGTGTCCCTGGTGCTGAGTGAGCTTCGGCCCAACGACTCCGGC
ATCTACCGCTGCGAGGTCCAGCACGGCATAGATGACAGCAGCGATGCCGTGGAGGTCAAGGTCAAAGgtg
agggggcaggacccaggaaggttcccccagggtgggagcccacagtgtgaggggggagcaaatgctctgga
aggcacccagtgagctagaaaaggagtgggtagagacagacctttgttacctcctttctctcttggggtg
gacagactcctgagcacagcctggggcaggcccctcagcagtagagagagcacaatgtgaactttaccct
cgtgtacaggggagtgccaaggagggtgagaggaggtcagggtggggtgatcatgcctcagggctcctct
ctgcccctcagGGGTCGTCTTTCTCTACCGGGAAGGCTCTGCCCGCTATGCTTTCTCTTTCGCTGGGGCC
CAGGAGGCCTGTGCCCGCATTGGAGCCCGCATTGCCACCCCGGAGCAGCTCTATGCCGCCTACCTCGGGG
GCTATGAACAGTGTGATGCTGGCTGGCTATCTGACCAGACCGTGAGgtgagcagggttgggacggggggat
cctgtggaccaggagcttctagttgtgtctggaagtggcagtgggccctggttctgccaggggctgctgg
```

Exon 1
5'UTR

Exon 2

Exon 3

Exon 4

Fig. 4. Position of the EFS microdeletion. Genomic DNA from canine chromosome 7 highlighting exons 1–4 of BCAN and the position of the 15.7 kb EFS microdeletion (grey shading).

This quantitative analysis also confirmed that EFS is associated with a loss of *BCAN* promoter/regulatory elements and exons 1-3 in heterozygous carriers and homozygous affected animals (Fig. 5).
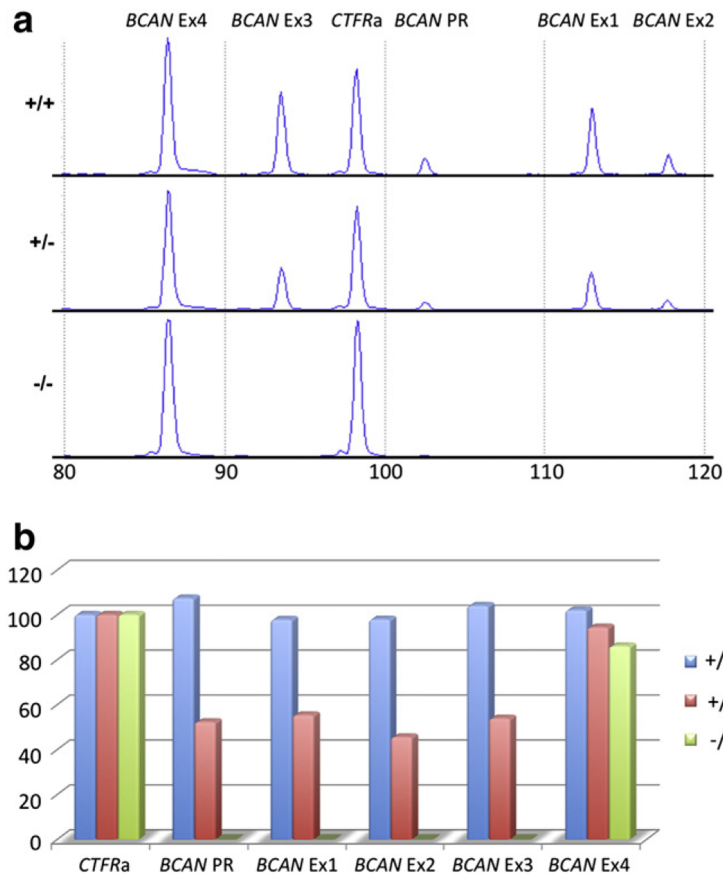
Fig. 5. Confirmation of the BCAN microdeletion using Multiplex Ligation-dependent Probe Amplification. (a) MLPA analysis revealed robust detection of a control probe (CFTRa) and probes for the BCAN promoter/regulatory region (PR) and exons 1–4. However, signals for probes PR and exons 1–3 were reduced by 46–55% in heterozygous (+/-) animals and abolished in homozygous animals (-/-), consistent with a loss of probe binding sites in genomic DNA.

Rapid genotyping using multiplex genotyping in different dog populations:

To assess the prevalence of the EFS microdeletion, we used our multiplex PCR assay to test several CKCS populations and other dog breeds (Table 1). All affected dogs in our study cohort - from the UK (n = 2), USA (n = 6) and New Zealand (n = 2), were homozygous for the *BCAN* microdeletion. In addition, all obligate carriers were heterozygous (n = 2 from the USA and n = 6 from New Zealand). In animals related to affected dogs, we found 9 normal animals and 10 carriers. Interestingly, in this group,

Table 1

BCAN genotypes in CKCS cohorts and other dog breeds.

Phenotype	Normal	Carrier	Affected
Study CKCS EFS affected	0/10	0/10	10/10
Study CKCS EFS carrier	0/8	8/8	0/8
Study CKCS related to affected or carrier	9/21	10/21	2/21
CKCS with no EFS history	135/155	20/155	0/155
54 dog breeds with no EFS history	93/93	0/93	0/93

Genotypes revealed by multiplex PCRs were determined as described in Materials and methods. Dogs were evaluated on the basis of available clinical data and placed into one of the phenotype categories above. Note that all clinically affected animals were homozygous for the BCAN deletion, whilst obligate carriers were heterozygous. As well as wild-type animals and carriers of the BCAN deletion, two dogs homozygous for the BCAN deletion, which were not reported to have classical clinical signs of EFS, were detected in a cohort of animals related to known EFS dogs. Carriers were also detected in CKCS with no history of EFS, but not in control DNA samples from 54 other dog breeds including: Airedale terrier, Akita Basenji, American Staffordshire Terrier, American Cocker Spaniel, American Eskimo Dog, Australian Shepherd, Akita Basenji, Bernese Mountain Dog, Bluetick Coonhound, Border collie, Boston Terrier, Boxer, Boykin Spaniel, Briard, Bull Mastiff, Bulldog, Cairn Terrier, Catahoula Leopard Dog, Chihuahua, collie, Dachshund, Dalmatian, English Setter, English Springer Spaniel, Flat Coated Retriever, German Shepherd, Giant Schnauzer, Golden Retriever, Great Dane, Havanese, Siberian Husky, Irish Setter, Italian Greyhound, Labrador Retriever, Miniature Pinscher, Miniature Poodle, Miniature Schnauzer, New Guinea Singing Dog, Norwegian Elkhound, Petit Basset Griffon Vendeen, Pomeranian, Portuguese Podengo Pequeno, Pug, Pyrenean Shepherd, Schipperke, Shetland Sheepdog, Swedish Vallhund, Tibetan Terrier, Toy Fox Terrier, Weimaraner, Welsh Terrier, West Highland White Terrier, Wire Fox Terrier and Yorkshire Terrier. Where possible, two unrelated dogs were tested for each breed.

two dogs without a classical clinical history of EFS were homozygous for the microdeletion. Lastly, in dogs with no known clinical history of EFS sourced from the USA, the carrier frequency was 12.9% (20/155) suggesting that the EFS microdeletion is present at a high frequency in this population. Notably, the mutation was not detected in multiplex PCRs conducted on control DNA samples from 54 other breeds of dog (Table 1).

DISCUSSION

The genomic architecture of pure-bred dog lines is ideal for the identification of loci responsible for autosomal recessive traits using genome-wide association mapping (14,21). In this study, we demonstrate that this technique can be used successfully on minimal samples sets, since we located the EFS locus using DNA samples from only five affected and seven breed-matched control dogs. Since a homozygous haplotype spanning 6.35 Mb was identified in affected animals, it is questionable whether further SNP typing would have generated additional useful data. In fact, a single recombination event in an obligate carrier allowed us to narrow the critical interval to 3.48 Mb. This region contained >100 genes, including ligand-gated ion channels, K+ channels, transporters, mitochondrial proteins and several genes known to be involved in neurological disorders in humans. For example, mutations in *CHRNB2* are associated with nocturnal frontal lobe epilepsy (22) and *TPM3* mutations are associated with nemaline myopathy (23). However, many of these genes were rapidly eliminated as candidates due to either: i) poor correlation of EFS clinical signs with the equivalent human disorders or ii) systematic resequencing of the genes. Consistent with the unique clinical signs observed in affected dogs, we discovered that a homozygous microdeletion affecting *BCAN* is associated with EFS in CKCS dogs, confirming that this disorder is inherited in an autosomal recessive manner. This mutation was not detected in control DNA samples from 54 other dog breeds, confirming the unique nature of this genomic rearrangement.

Brevican belongs to the lectican family of aggregating extracellular matrix (ECM)

41

proteoglycans, which comprises aggrecan, brevican, neurocan and versican. Although

mutations in the aggrecan and versican genes (*ACAN*: 15q26.1 and *VCAN*: 5q14.2-14.3)

have been linked to different connective tissue disorders (24-26; Fig. 6), no mutations in

the brevican or neurocan genes (*BCAN*: 1q23.1 and *NCAN*: 19p13.11) have been

identified to date. Brevican and neurocan are highly expressed in the central nervous

system, where they are found in specialized extracellular matrix structures called

perineuronal nets that play a role in cell adhesion, migration, axon guidance and neuronal

plasticity (27).
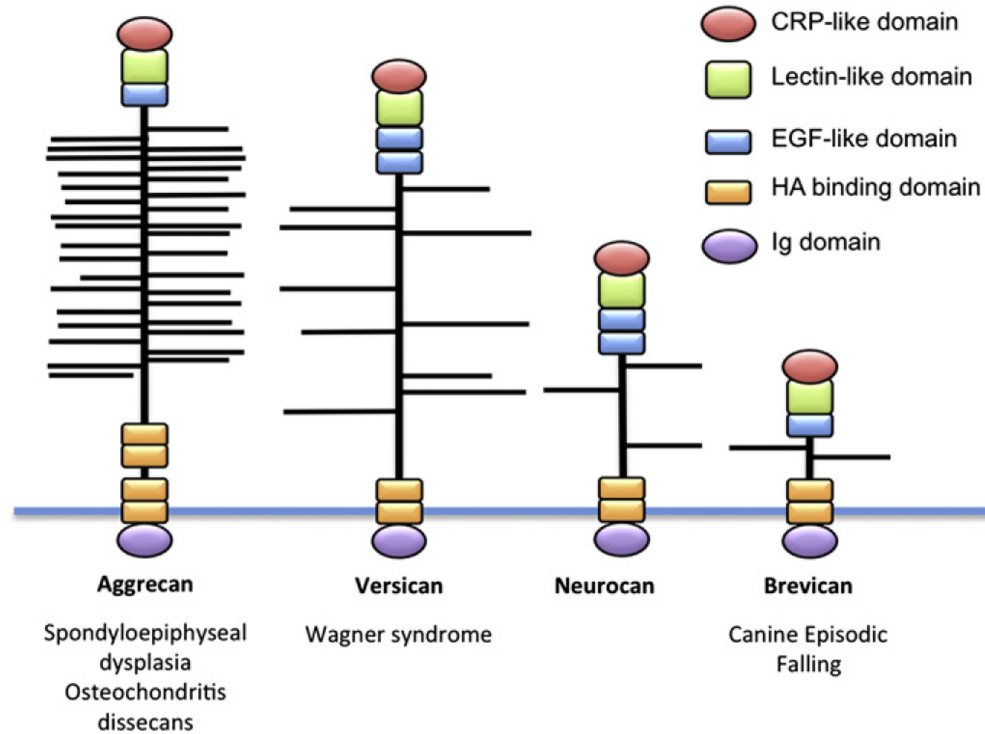
Fig. 6. Modular organization of the superfamily of hyaluronan-binding proteins and associated disorders. Mutations in ACAN, encoding aggrecan—a major component of cartilage, have been implicated in spondyloepiphyseal dysplasia type Kimberley (24) and familial osteochondritis dissecans (25). Mutations in VCAN, encoding versican, are associated with Wagner syndrome and erosive vitreoretinopathy, disorders affecting the connective tissue of the eye (26). Modified from (28).

Brevican, versican, HAPLN2/BRAL1, tenascin-R and phosphacan are also present at the nodes of Ranvier on large diameter myelinated axons (18,29) where cations are accumulated and depleted in the local extracellular nodal region during action potential propagation. The ECM complex at nodes of Ranvier is thought to play a pivotal role in maintaining a local microenvironment, acting as a diffusion barrier for K+ and Na+ around the perinodal extracellular space (17,18). Thus, disruption of ECM complexes governing synapse stability and nerve conduction velocity are likely to underlie the EFS phenotype. Certainly, since EFS appears to result from a central nervous system rather

43

than a muscle defect, the associated sarcoplasmic reticulum pathology is likely to be a secondary manifestation of muscle overstimulation (30). Interestingly, *BCAN* knockout mice were not reported (27) to have a phenotype similar to EFS - although an increased grip-strength was noted, which could be indicative of increased muscle tone. However, it is questionable whether current mouse phenotyping tests will reveal neurological disorders evoked by strenuous exercise, stress, apprehension or excitement, since these conditions are generally avoided in mouse care. It is also noteworthy that EFS clearly shows variable age of onset and penetrance. For instance, we identified two dogs that were homozygous for the *BCAN* microdeletion that had not shown classical clinical signs of EFS in the presence of their owners. Interestingly, similar findings were reported for the dynamin (*DNM1*) mutation (p.R256L) underlying exercise-induced collapse in Labrador retrievers (31). It is plausible that these dogs have not been exposed to sufficient exercise or excitement to trigger an episode, or that genetic or environmental modifying factors (e.g. diet) affect the onset of these disorders. However, it is notable that one of the two asymptomatic dogs in our study was described as 'exercise resistant' by their owner, suggesting that adaptive behavior may occur in dogs that are homozygous for the *BCAN* microdeletion. It must also be emphasized that such non-classical cases are rare - we did not find a single dog that was homozygous for the *BCAN* microdeletion in 155 CKCS with no clinical history of EFS (Table 1).

In summary, we have shown that an inactivating microdeletion affecting *BCAN* is associated with EFS in CKCS dogs. Identification of the deletion breakpoint has allowed

the development of diagnostic tests that have revealed a high prevalence of carriers (12.9%) in clinically unaffected dogs from the USA. These genetic tests (available via Laboklin: http://www.laboklin.co.uk/) will enable future identification of heterozygous animals (which have no discernable phenotypic difference to wild-type animals) allowing directed breeding programs to be implemented, and confirmatory diagnosis and appropriate pharmacotherapy of affected animals. Since this also represents the first report of a genetic disorder involving a neuronal-specific ECM proteoglycan, we suggest that *BCAN* and *NCAN* should be considered as candidate genes for genetic analysis in unresolved cases of human disorders with similar clinical presentations to EFS, such as paroxysmal exercise-induced dyskinesias (32) or episodic ataxias (12).

REFERENCES

1. Herrtage ME, Palmer AC (1983) Episodic falling in the cavalier King Charles spaniel. Vet Rec 112:458-459.

2. Wright JA, Brownlie SE, Smyth JB, Jones DG, Wotton P (1986) Muscle hypertonicity in the cavalier King Charles spaniel-myopathic features. Vet Rec 118:511-512.

3. Wright JA, Smyth JB, Brownlie SE, Robins M (1987) A myopathy associated with muscle hypertonicity in the Cavalier King Charles spaniel. J Comp Pathol 97:559-565.

4. Rusbridge C (2005) Neurological diseases of the Cavalier King Charles spaniel. J Small Anim Pract 46:265-272.

5. Shiang R, Ryan SG, Zhu YZ, Hahn AF, O'Connell P et al. (1993) Mutations in the α1 subunit of the inhibitory glycine receptor cause the dominant neurologic disorder, hyperekplexia. Nat Genet 5:351-358.

6. Rees MI, Lewis TM, Kwok JB, Mortier GR, Govaert P et al. (2002) Hyperekplexia associated with compound heterozygote mutations in the β-subunit of the human inhibitory glycine receptor (*GLRB*). Hum Mol Genet 11:853-860.

7. Rees MI, Harvey K, Pearce BR, Chung SK, Duguid IC et al. (2006) Mutations in the gene encoding GlyT2 (*SLC6A5*) define a presynaptic component of human startle disease. Nat Genet 38:801-806.

8. Harvey RJ, Topf M, Harvey K, Rees MI (2008) The genetics of hyperekplexia: more than startle! Trends Genet 24:439-447.

9. Gill JL, Capper D, Vanbellinghen JF, Chung SK, Higgins RJ et al. (2011) Startle disease in Irish wolfhounds associated with a microdeletion in the glycine transporter GlyT2 gene. Neurobiol Dis 43:184-189.

10. Garosi LS, Platt SR, Shelton GD (2002) Hypertonicity in Cavalier King Charles spaniels. J Vet Intern Med 16:330.

11. Thomas RH, Stephenson JBP, Harvey RJ, Rees MI (2010) Hyperekplexia: Stiffness, startle and syncope. J Ped Neurol 8:11-14.

12. Tomlinson SE, Hanna MG, Kullmann DM, Tan SV, Burke D (2009) Clinical neurophysiology of the episodic ataxias: insights into ion channel dysfunction *in vivo*. Clin Neurophysiol. 120:1768-1776.

13. Matthews E, Hanna MG (2010) Muscle channelopathies: does the predicted channel gating pore offer new treatment insights for hypokalaemic periodic paralysis? J Physiol 588:1879-1886.

14. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC et al. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet 39:1321-1328.

15. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575.

16. Hirakawa S, Oohashi T, Su WD, Yoshioka H, Murakami T et al. (2000) The brain link protein-1 (BRAL1): cDNA cloning, genomic structure, and characterization as a novel link protein expressed in adult brain. Biochem Biophys Res Commun 276:982-

989.

17. Oohashi T, Hirakawa S, Bekku Y, Rauch U, Zimmermann DR et al. (2002) Bral1, a brain- specific link protein, colocalizing with the versican V2 isoform at the nodes of Ranvier in developing and adult mouse central nervous systems. Mol Cell Neurosci 19:43- 57.

18. Bekku Y, Vargová L, Goto Y, Vorísek I, Dmytrenko L et al. (2010) Bral1: its role in diffusion barrier formation and conduction velocity in the CNS. J Neurosci 30:3113- 3123.

19. Chuzhanova N, Abeysinghe S, Krawczak M, Cooper D (2003) Translocation and gross deletion breakpoints in human inherited disease and cancer II: Potential involvement of repetitive sequence elements in secondary structure formation between DNA ends. Hum Mut 22:245-251.

20. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F et al. (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30:e57.

21. Drögemüller C, Becker D, Brunner A, Haase B, Kircher P et al. (2009) A missense mutation in the SERPINH1 gene in Dachshunds with osteogenesis imperfecta. PLoS Genet 5:e1000579.

22. De Fusco M, Becchetti A, Patrignani A, Annesi G, Gambardella A et al. (2000) The nicotinic receptor β2 subunit is mutant in nocturnal frontal lobe epilepsy. Nat Genet 26:275-276.

23. Laing NG, Wilton SD, Akkari PA, Dorosz S, Boundy K et al. (1995) A mutation in

the alpha tropomyosin gene *TPM3* associated with autosomal dominant nemaline myopathy NEM1. Nat Genet 9:75-79.

24. Gleghorn L, Ramesar R, Beighton P, Wallis G (2005) A mutation in the variable repeat region of the aggrecan gene (*AGC1*) causes a form of spondyloepiphyseal dysplasia associated with severe, premature osteoarthritis. Am J Hum Genet 77:484-490.

25. Stattin EL, Wiklund F, Lindblom K, Onnerfjord P, Jonsson BA, et al. (2010) A missense mutation in the aggrecan C-type lectin domain disrupts extracellular matrix interactions and causes dominant familial osteochondritis dissecans. Am J Hum Genet 86:126-137.

26. Miyamoto T, Inoue H, Sakamoto Y, Kudo E, Naito T et al. (2005) Identification of a novel splice site mutation of the CSPG2 gene in a Japanese family with Wagner syndrome. Invest Ophthalmol Vis Sci 46:2726-2735.

27. Brakebusch C, Seidenbecher CI, Asztely F, Rauch U, Matthies H et al. (2002) Brevican- deficient mice display impaired hippocampal CA1 long-term potentiation but show no obvious deficits in learning and memory. Mol Cell Biol 22:7417-7427.

28. Maeda N, Fukazawa N, Ishii M (2010) Chondroitin sulfate proteoglycans in neural development and plasticity. Front Biosci 15:626-644.

29. Bekku Y, Rauch U, Ninomiya Y, Oohashi T (2009) Brevican distinctively assembles extracellular components at the large diameter nodes of Ranvier in the CNS. J Neurochem 108:1266-1276.

30. Engel AG, Banker BQ (2004) Ultrastructural changes in diseased muscle. In: Engel

AG, Franzini- Armstrong CF (eds.). Myology 3rd ed. McGraw-Hill, New York, pp. 749-887.

31. Patterson EE, Minor KM, Tchernatynskaia AV, Taylor SM, Shelton GD et al. (2008) A canine *DNM1* mutation is highly associated with the syndrome of exercise-induced collapse. Nat Genet 40:1235-1239.

32. Weber YG, Lerche H (2009) Genetics of paroxysmal dyskinesias. Curr Neurol Neurosci Rep 9:206–211.

CHAPTER III

GENOME-WIDE ASSOCIATION STUDIES FOR MULTIPLE DISEASES OF THE

GERMAN SHEPHERD DOG

Kate L. Tsai[a], Rooksana E. Noorai[a], Alison N. Starr-Moss[a], Pascale Quignon[b,c], Caitlin J. Rintz[a], Elaine A. Ostrander[b], Jörg M. Steiner[d], Keith E. Murphy[a], Leigh Anne Clark[a]

[a] Department of Genetics and Biochemistry, College of Agriculture, Forestry and Life Sciences, Clemson University, Clemson, SC 29634, USA

[b] Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

[c] Institut de Génétique et Dévelopment de Rennes, CNRS-UMR6062, Université de Rennes 1, 2 avenue Prof. Léon Bernard, CS34317, Rennes Cedex 35043, France

[d] Department of Small Animal Medicine and Surgery, College of Veterinary Medicine and Biological Sciences, Texas A&M University, College Station, TX 77843, USA

ABSTRACT

The German shepherd dog (GSD) is an extremely popular working and companion breed for which over 50 hereditary diseases are documented. Herein, we generated SNP profiles for 197 GSDs using the Affymetrix v2 canine SNP array and employed a genome-wide association strategy to find loci associated with four diseases: pituitary dwarfism, degenerative myelopathy (DM), congenital megaesophagus (ME), and pancreatic acinar atrophy (PAA). A strongly associated locus on chromosome 9 was detected for pituitary dwarfism, and is proximal to a plausible candidate gene, *LHX3*. Results for DM confirm a major locus encompassing *SOD1*, in which an associated point mutation was previously identified, but do not suggest significant modifier loci. Several SNPs on chromosome 12 were found to be associated with ME and a 4.7 Mb haplotype block was identified. Analysis of additional affected dogs for a SNP within the haplotype provides further support for the association. Results for PAA indicate more complex genetic underpinnings. Loci on 23 chromosomes reached genome-wide significance. No major locus was detected and only two weakly associated haplotype blocks, on chromosomes 7 and 12, could be detected. These data suggest that PAA may be governed by multiple loci with small effects, or may be a heterogeneous disorder.

52

INTRODUCTION

Shepherds and hounds were the earliest dog breeds developed to serve in specialized roles (1). The German shepherd dog (GSD), for which the standard was originally developed in 1899, was bred for utility and intelligence and to be a multi-purpose servant of humans (1). Aptitude, temperament, structural efficiency and other natural skills were selected for over aesthetic traits. Today, the GSD is among the most common breeds trained in specialized jobs for the military, law enforcement, and service/assistance programs (2). The GSD is also an extremely popular companion, ranking second in breed registration statistics reported by the American Kennel Club in 2010. Unfortunately, despite a large population size and outbreeding practices, over 50 hereditary diseases plague the GSD (3).

Pituitary dwarfism is a rare disease of GSDs characterized by an abnormally small body structure due to a deficiency in pituitary growth hormone. Dogs appear to be normal at birth, but signs of slowed growth are usually evident by 2-3 months of age. There are no treatments available for dogs and their lifespan may be significantly shortened. Pituitary dwarfism in the GSD is inherited in an autosomal recessive fashion (4). Combined pituitary hormone deficiency (CPHD) in the human is a type of growth hormone deficiency that results from a decrease in several hormones necessary for pituitary development. Mutations that cause congenital CPHD in the human have been identified in several transcription factors important for pituitary development, such as

PIT1, *PROP1*, *LHX3*, *LHX4*, *HESX1* (5). Sequencing of *PIT1* (6), *PROP* (7)*, LHX4* (8), and *LIFR* (9), did not reveal any causative variants.

Degenerative myelopathy (DM) is a late-onset neurodegenerative disease characterized by ataxia and weakness in the hind limbs. Symptoms of DM worsen over time, either steadily or in phases, and eventually result in complete paraplegia. In some cases, DM may progress up the spinal cord and cause forelimb weakness and respiration difficulties (10). Age of onset is generally between 5 and 14 years, with a mean onset of 9 years (11). The clinical signs observed in DM are general indicators of spinal cord disorders; definitive diagnosis for DM can only be made by histopathological examination of spinal cord tissue postmortem (12). Most owners of affected dogs elect euthanasia within several months of the onset of clinical signs (13). DM is most prevalent in the GSD (12). A recessive missense mutation in *SOD1* is associated with DM in several breeds, including GSDs (14). DM has been proposed to be a large-animal model for amyotrophic lateral sclerosis (ALS). Approximately 20% of hereditary ALS cases are attributed to mutation of *SOD1* (15).

Congenital idiopathic megaesophagus (ME) is characterized by dilation and hypomotility of the esophagus. The result of these anomalies is regurgitation several minutes to hours after eating. Regurgitation episodes may occur as often as several times a day or as infrequently as once every few days. Symptoms for congenital ME typically begin around five weeks of age after weaning onto solid food. Affected puppies are

54

malnourished, show a general failure to thrive, and are at risk for aspiration pneumonia. Diagnosis of ME is achieved by standard X-rays and/or fluoroscopy (barium swallow) (16). ME is typically treated by feeding a high calorie, liquid diet from a raised dish (16). Mortality is high in affected neonates. Many survivors require lifelong management of the disorder, but other cases resolve by four to six months of age (17). ME is prevalent in the GSD, Great Dane, miniature schnauzer, Rhodesian ridgeback, dachshund, and wire fox terrier breeds, and is also reported to occur at lower frequencies in many other breeds (18). Several loci are thought to be responsible for ME in the dog (17), but the mode of inheritance in the GSD has not been investigated.

Pancreatic acinar atrophy (PAA) is an autoimmune disease characterized by the selective atrophy of the acinar cells of the pancreas, which produce and secrete digestive enzymes (19). PAA is the most common cause of exocrine pancreatic insufficiency (EPI) in the dog (20) and is diagnosed through histopatholgic examination (21). EPI is diagnosed through measurement of serum canine trypsin-like immunoreactivity (cTLI), with concentrations < 2.5 μg/L diagnostic for EPI (22). Ninety-six percent of affected dogs present with clinical signs of EPI, including weight loss, increased appetite, and soft stools, by five years of age (20,23). Affected dogs can be treated with pancreatic enzyme supplements, although dogs with EPI are often euthanized because of the expense of treatment or poor treatment response (24,25). PAA is a disorder that is widespread in the GSD population, but also affects many other breeds including collies (26,27). Recent inheritance studies in the GSD show that EPI is likely a polygenic disorder (28).

55

The development of high-throughput genotyping technologies has facilitated the discovery of genes responsible for simple and complex phenotypes of the dog (29,30). The population structure of the dog is particularly well-suited for association mapping techniques because genetic bottlenecks during breed formation created large blocks of linkage disequilibrium within breeds (29). Reported here is the use of a large cohort of GSDs to investigate the afore- mentioned four diseases that segregate in the population. Genetic profiles for 197 GSDs were generated using the Affymetrix v2 canine SNP array. Using a single data set, we carried out genome-wide association studies (GWAS) to identify risk loci for four diseases of the GSD.

MATERIALS AND METHODS

Sample Collection

Samples were recruited from purebred GSDs that had been diagnosed with EPI, DM, or ME through standard veterinary diagnostic procedures. DM cases were diagnosed by veterinary neurologists. Dogs were considered to have congenital idiopathic ME if they did not have conditions that cause secondary ME, such as persistent right aortic arch and myasthenia gravis. Purebred GSDs collected as controls were at least six years of age and, to the owners' knowledge, had no immediate family members affected with EPI, DM, or ME. Cases and controls were selected to have as few close relatives (within three generations) in the study cohort as possible. Samples were acquired from the United States, Canada, Germany, and the Netherlands. Owners were asked to report all known hereditary conditions, as well as coat color phenotypes.

Whole blood or buccal swabs were submitted by dog owners. Blood samples were collected by owners' private veterinarians. All samples were collected in accordance with protocols approved by the Clinical Research Review Committee (CRRC No. 07-04) at Texas A&M University and the Clemson University Institutional Review Board (IBC2008-17). Genomic DNA was isolated using the Gentra Puregene Blood Kit (Qiagen, Valencia, CA, USA). Serum samples were collected from all EPI cases and controls and submitted to the Gastrointestinal Laboratory at Texas A&M University for measurement of cTLI to confirm clinical status. Concentrations ≤2.5 μg/L were considered diagnostic for EPI. Control dogs had concentrations ≥5.7 μg/l. All EPI cases were assumed to result from PAA, but tissue samples were not available to differentiate EPI due to PAA or another cause of EPI.

GWAS

SNP genotypes were generated using the Affymetrix Canine SNP Array version 2 (http://www.broadinstitute.org/mammals/dog/caninearrayfaq.html). Arrays were processed at the NHGRI or Clemson University Genomics Institute (http://www.genome.clemson.edu/) using the GeneChip human mapping 250K Sty assay kit (Affymetrix, Santa Clara, USA). The GeneChip human mapping 500K assay protocol was followed, but with a hybridization volume of 125 μl (29). Raw CEL files were genotyped using Affymetrix Power Tools software. SNPs having >10% missing data and ≥60% heterozygosity were removed from the data set. Case/control analyses were performed on the 48,415 SNPs from the ''platinum'' set using PLINK (31), with P_{raw} <

0.0001 considered significant. To determine P_{genome} values (EMP2), permutation testing (100,000) was performed for PAA, DM, and ME analyses using PLINK. A less conservative correction, Benjamini–Hochberg, was also performed (32). Analysis of population stratification in 197 GSDs was conducted using identity-by-state clustering and multidimensional scaling in PLINK. The data, in a reduced representation, were plotted in two dimensions: principal components 1 and 2.

Genotyping of SNP 12.60274687

Primers flanking the SNP 12.60274687 were designed: forward: 5$^{'}$-TGAGCACACAGAGGTGAGACAT-3$^{'}$ and reverse: 5$^{'}$-CAGTGGGAGGGTTTAGGAAGAGAT-3$^{'}$. SNP 12.60274687 was amplified for 35 additional ME dogs using Thermo ReddyMix (Thermo Fisher Scientific, Waltham, MA, USA) following the recommended proto- col. Thermal cycling conditions were as follows: 95°C for 15 min; 14 cycles of 95°C for 30 s, 62°C for 1 min (decreasing 0.5°C each cycle), 72°C for 1 min; 20 cycles of 95°C for 30 s, 55°C for 1 min, 72°C for 1 min; 72°C for 10 min. PCR amplicons were analyzed by electrophoresis on agarose gels. Products were purified by adding 0.5 unit exonuclease I (New England Biolabs, Ipswich, MA, USA) and 0.25 unit of shrimp alkaline phosphatase (Promega, Madison, WI, USA) and incubating for 30 min at 37°C followed by 20 min at 80°C. Nucleotide sequencing was accomplished using the Big Dye Terminator version 3.1 Cycle Sequencing kit (Applied Biosystems, Carlsbad, CA, USA). Products were purified using Spin-50 mini columns with water (BioMax, Inc., Arnold, MD, USA) and resolved on an ABI Prism

58

3730 Genetic Analyzer (Applied Bio- systems). Sequences were analyzed to determine SNP genotypes. Odds ratios (OR), critical intervals, and x^2 tests were calculated to assess the differences in genotype frequencies between the two populations using 2 x 2 contingency tables.

RESULTS

Validation

In total, 197 GSD samples were acquired from the United States, Canada, Germany, and the Netherlands. Two GSDs were affected with EPI and DM. One GSD was affected with EPI and ME. Principal component analysis with PLINK showed no evidence for population stratification within our GSD cohort (n = 197) using all ''platinum'' SNPs (Fig. 1). To validate both our genotype data and statistical approach, we first mapped a positive control trait. Ten dogs in the study cohort had white coats, a recessive trait caused by the e allele of MC1R in the GSD (33). GWAS using 197 GSDs (10 cases, 187 controls) reveals a strongly associated region encompassing MC1R on chromosome 5 (Fig. 2a). SNP 5.66900853, the most significant result (P_{raw} = 3.61 9 10^{-22}), is located approximately 208 kb adjacent to the E locus. The lack of subpopulations, combined with our ability to accurately map a known locus, indicates the utility of this data set.
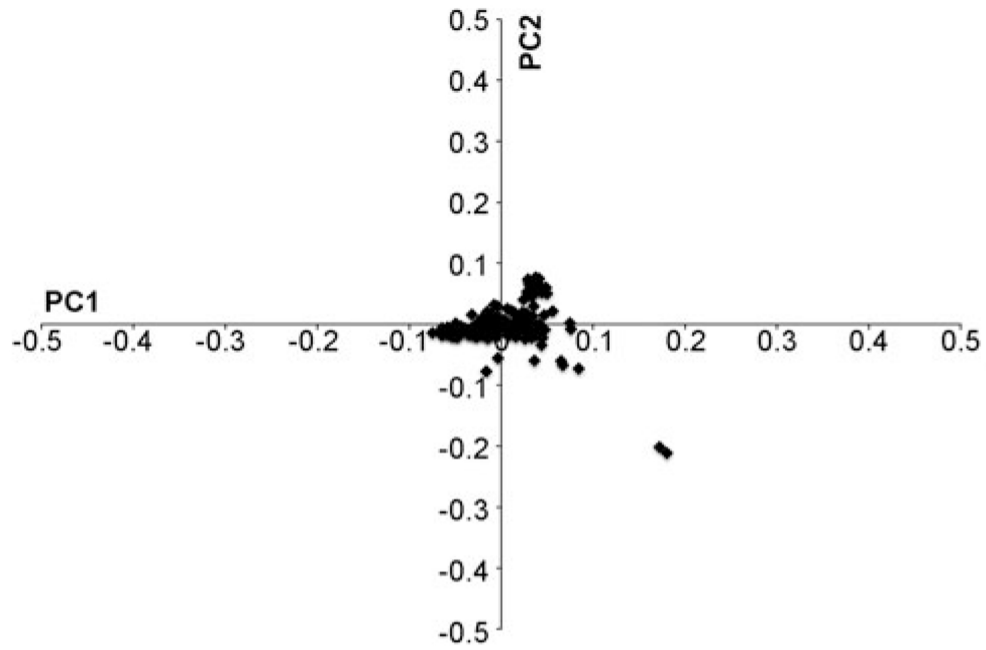
Fig.1 Principal component analysis of the GSD cohort shows all 197 dogs clustering together. The *x* axis is principal component 1 and the y axis is principal component 2. Two individuals that appear apart from the main grouping have the largest proportion of missing data (22.7 and 22.4% overall, compared to the average missing data rate of 12.6%)
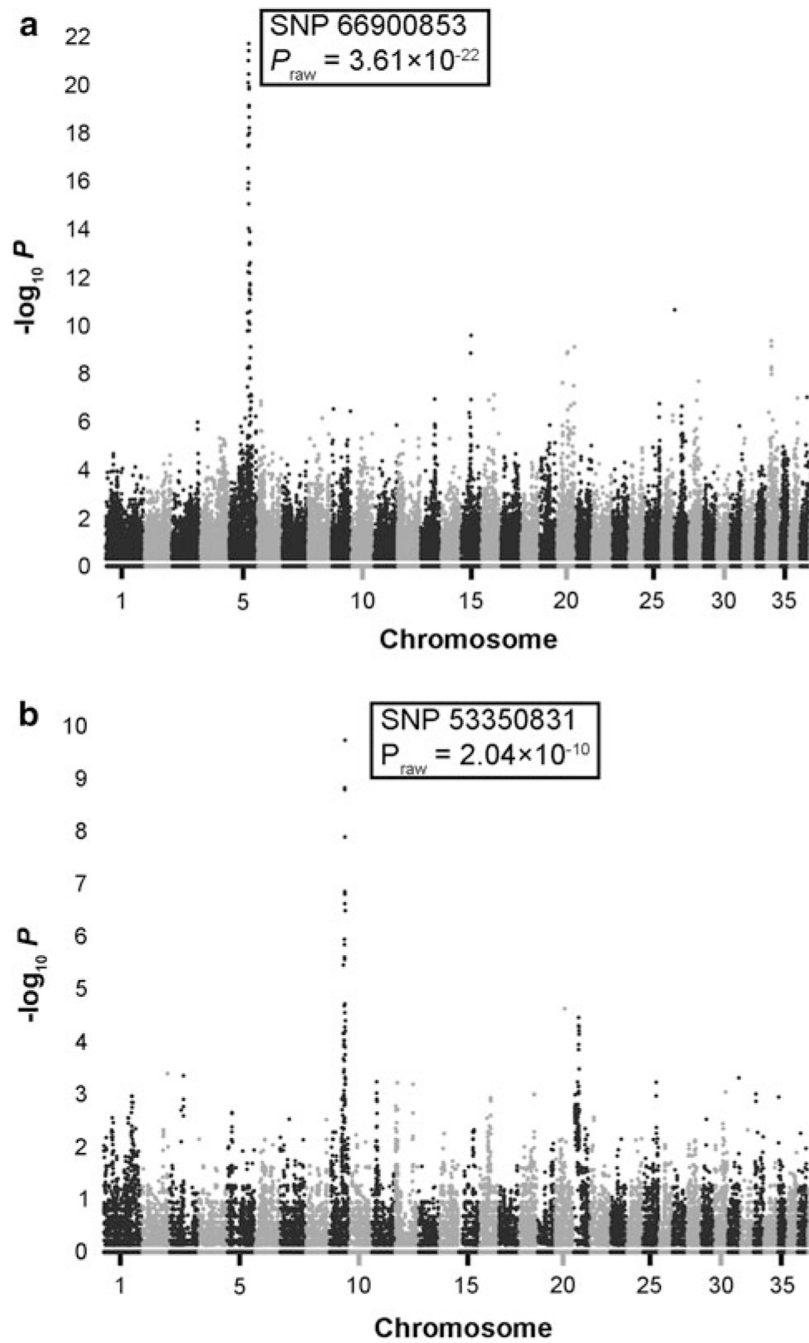
Fig. 2 Manhattan plots showing the results for GWAS using 48,415 SNPs. The genome-wide P values (-log$_{10}$ *P*) for each SNP are plotted against position on each chromosome. **a** Ten white GSDs versus 187 nonwhite GSDs show a strong association on chromosome 5. **b** Four pituitary dwarf GSDs versus 193 normal-sized GSDs show a strong association on chromosome 9.

61

Pituitary dwarfism

Four dogs in our control population have pituitary dwarfism. GWAS using these four cases and 193 normal-sized controls, showed a major region of association on chromosome 9 (Fig. 2b). Within this region, all four pituitary dwarfs were homozygous for a haplotype extending from SNP 9.52031784 to SNP 9.54102643. The haplotype was never observed in the homozygous state among the control dogs, and was heterozygous in 15 control dogs. The most significant result, SNP 9.53350831 (P_{raw} = 2.04 × 10^{-10}), is located approximately 764 kb from a candidate gene, *LHX3*.

Degenerative myelopathy

GWAS with 100K permutations for DM were carried out using all cases collected (n = 15) and control dogs that were at least eight years of age and had exhibited no clinical signs of DM (n = 69). The two most significant results were for SNPs 31.28909487 (P_{raw} = 8.51 × 10^{-06}, P_{genome} = 0.1461) and 31.30528862 (P_{raw} = 5.59 × 10^{-05}, P_{genome} = 0.5771) (Fig. 3a). These SNPs on chromosome 31 are located 650 kb upstream and 965 kb downstream of *SOD1*, respectively. Evaluation of genotypes in this region showed that only four of the 15 cases were homozygous at both loci. Five cases were homozygous for the associated allele of SNP 31.30528862, while nine cases were homozygous for the associated allele of SNP 31.28909487. In the control population, six of 69 GSDs were also homozygous for the same allele of SNP 31.28909487. A third significant result was detected on chromosome 10 for SNP 10.64801770 (P_{raw} = 6.06 × 10^{-05}, P_{genome} = 0.6066).
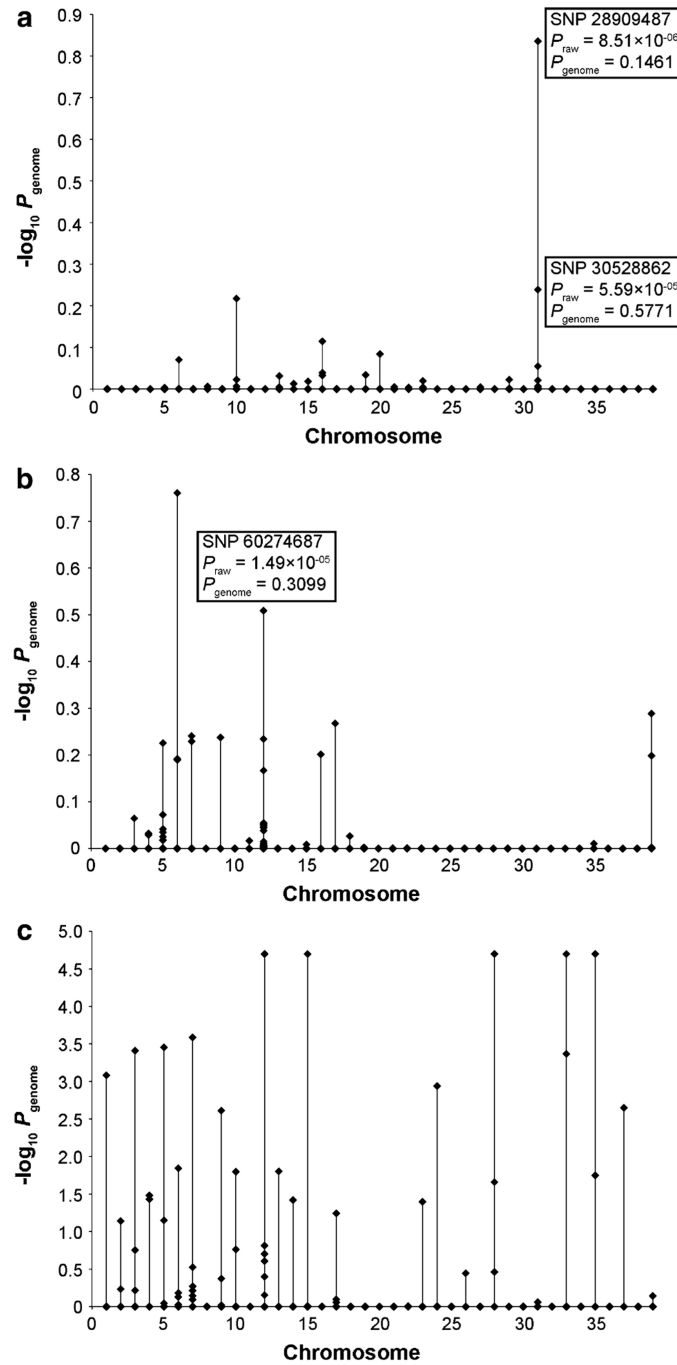
Fig. 3 GWAS using 48,415 SNPs with 100,000 permutations. The genome-wide adjusted P values ($-\log_{10}$ *P*) for each SNP are plotted by chromosome. **a** Fifteen GSDs with DM versus 69 healthy control GSDs. **b** Nineteen GSDs with ME versus 177 healthy control GSDs. c One-hundred GSDs with PAA versus 79 healthy control GSDs.

Megaesophagus

GWAS with 100,000 permutations were carried out using 19 cases and 177 control dogs. Significant results were obtained for 14 SNPs on eight different chromosomes (Fig. 3b). Of the 14 significant results, three were located within an eight Mb region on chromosome 12. Five other SNPs within this region are approaching significance. The most significant of these is SNP 12.60274687 ($P_{raw} = 1.49 \times 10^{-5}$, $P_{genome} = 0.3099$). Analysis of genotypes for this SNP revealed that all affected dogs share a common allele (frequency = 0.83) that was less prevalent among control dogs (frequency = 0.46). To further investigate this finding, 35 additional ME GSDs were collected and genotyped for the SNP. Genotype frequencies and odds ratios for SNP 12.60274687 are shown in Table 1. Further analysis of genotypes on chromosome 12 revealed a haplotype block extending 4.7 Mb that is present in all affected dogs.

Table 1 Frequency data for SNP 12.60274687 in GSDs with and without ME

	G/G[b] (%)	G/A (%)	A/A (%)
ME$_{array}$ ($n = 19$)	12 (63)	7 (37)	0 (0)
ME$_{all}$ ($n = 54$)	33 (61)	20 (37)	1 (2)
Control ($n = 172$[a])	36 (21)	84 (49)	52 (30)

Observed genotypes and genotypic frequencies are presented for affected and control dogs. ME$_{array}$ = genotype frequencies of the 19 ME dogs run on Affymetrix SNP arrays; ME$_{all}$ = All ME dogs collected

[a] Four control dogs were N/N for SNP 12.60274687 on the array and were excluded from this analysis

[b] Genotype (G/G): $P = 6.84 \times 10^{-8}$, OR = 5.937 (95% CI = 3.071–11.475); Allele (G): $P = 2.77 \times 10^{-10}$, OR = 4.711 (95% CI = 2.817–7.878)

Pancreatic acinar atrophy

GWAS with 100,000 permutations were carried out using 100 cases and 79 control dogs (Fig. 3c). Significant P_{raw} values were obtained for 50 SNPs on 23 chromosomes (Supplementary Table 1). Chromosome 12 harbors seven significant SNPs,

64

three of which are found in a cluster: 12.3781476, 12.3845215, and 12.4698937.
Chromosome 7 harbors six significant SNPs, four are found in a cluster: 7.11222056,
7.12305239, 7.12310223, 7.12424701. On ten chromosomes, only one SNP each is
strongly linked with PAA. On the remaining 11 chromosomes, no two strongly associated
SNPs were located within 5 Mb of each other and no common haplotypes are detected.
Six SNPs had minor allele frequencies (MAF) < 15%.

DISCUSSION

Pituitary dwarfism

Autosomal recessive traits of the dog can be mapped using remarkably small
numbers of cases (34,35). In this study, a mere four cases were sufficient to detect strong
linkage to chromosome 9 and define a two Mb pituitary dwarfism haplotype. While the
critical interval harbors more than 40 loci, one gene is an intriguing candidate: *LHX3*, a
LIM homeodomain transcription factor vital for pituitary and motor neuron development
(36). Genetic variations within *LHX3* cause CPHD in humans (37,38). Although some
mutations in *LHX3* are also associated with a rigid cervical spine, mental retardation, and
deafness (37,39,40), the GSD pituitary dwarfs do not have clinical signs that indicate
nervous system involvement (41). In mice, complete loss of *LHX3* is lethal (42), but
mutations within the gene cause a dwarfism phenotype (43,44).

Recently, Voorbij et al. announced the availability of a genetic test for pituitary
dwarfism in the GSD (45), but information regarding the gene or mutation has not been

published to date. Dwarfism in the GSD is reportedly rare and utilization of the genetic test could facilitate rapid elimination of the phenotype. Our data are surprising because the frequency of the two Mb haplotype in our control dogs is four percent, suggesting a roughly eight percent carrier frequency in the general population. Moreover, this estimate is likely low because only dogs having the complete two Mb haplotype were considered to be carriers, thereby excluding dogs with partial overlapping haplotypes that may have the causative mutation.

Degenerative Myelopathy

DM is a devastating disorder of GDSs that has been difficult to eradicate from the breed, in part because of a late age of onset and poor diagnostic tools. We identified two SNPs on chromosome 31 that were strongly associated with the DM phenotype. The SNPs flank *SOD1*, the gene in which Awano et al. (14) identified an E40K missense mutation in DM-affected dogs from several breeds. The authors propose that the mutation causes DM in an autosomal recessive fashion (14). In our study, only nine of 15 cases were homozygous for the more tightly linked SNP, suggesting that either our relaxed diagnostic criteria (dogs were not diagnosed postmortem) may have led to the inclusion of dogs that were misdiagnosed, or that heterozygosity for the mutation can produce an affected phenotype. Recombination between the SNP, located 652 kb from the *SOD1* exon two mutation, may also be a factor (five cases were heterozygous). Only six of 69 dogs in our control population were homozygous for the SNP, corresponding to nine percent affected controls. Similarly, Awano et al. (14) reported that six percent of GSD

66

breed controls were homozygous for the *SOD1* mutation. The high frequency of clinically normal, aged controls that are homozygous for the mutation suggests the involvement of modifier loci and/or environmental factors. In our analysis, a second significant locus is detected on chromosome 10 and is supported by six proximal SNPs with *P* values approaching significance (< 0.00086). Awano et al. (14) detected weaker signals of association on chromosomes 6, 18, 20, and 35. Our data do not support regions of association on chromosomes 18 and 35, but do show modest signals on chromosomes 6 and 20.

Megaesophagus

The mode of inheritance of ME has not been previously investigated in the GSD, but it has been suggested that the frequent occurrence of affected puppies born to clinically normal parents serves as evidence for a recessive mutation. In this work we successfully mapped two simple autosomal recessive traits with ten and four cases and a complex recessive trait with 15 cases. GWAS using 19 dogs with ME, however, did not yield a single, strong region of association. Rather, minor associations were detected on eight different chromosomes. It is, therefore, unlikely that ME is inherited in an autosomal recessive fashion in the GSD. Similarly, evaluation of inheritance patterns of ME in the Miniature Schnauzer did not suggest a simple recessive mode of inheritance (17).

A locus on chromosome 12 is the only one to reach genome-wide significance and have supporting proximal SNPs (8 SNPs having *Praw* values <0.00014). To verify an association, 35 additional ME GSDs were collected and genotyped for the most strongly associated SNP in this region. In total, 53 of 54 ME cases were heterozygous or homozygous for the associated allele, which occurs at a significantly lower frequency in the control population. These data are consistent with an autosomal dominant mode of inheritance, with incomplete penetrance and/or modifying loci. On the other hand, it is also conceivable that the allele is fully penetrant but that the control population includes dogs with undetected ME. ME varies in severity and can resolve completely after a few months, making detection of mild cases difficult.

The dilation and lack of muscle contraction in the esophagus is thought to be a neurological defect rather than a muscular defect (16). The associated haplotype on chromosome 12 harbors approximately 12 genes, several of which are expressed in the nervous system. Additional genome-wide SNP analyses in ME GSDs are necessary to confirm this association.

Pancreatic acinar atrophy

Fifty SNPs, representing 45 loci, exceeded genome-wide significance thresholds for association with PAA. A majority of significant SNPs map to chromosome 12, although only three SNPs on the chromosome support the same locus. These SNPs map to the dog leukocyte antigen (DLA) complex in the centromeric region of the

chromosome. The association of this region with PAA is consistent with the autoimmune component of the disease. Previous work, we showed that pancreases from GSDs with PAA have increased expression of *DLA-88* (46). *DLA-88* is the most highly polymorphic class I locus of the DLA complex. DLA class II loci are associated with diabetes mellitus, anal furunculosis, hypoadrenocorticism, necrotizing meningoencephalitis, and several other autoimmune disorders of the dog (47-50). Canine systemic lupus erythematosus is also reported to be associated with alleles of the DLA class II loci, and GWAS revealed five additional associated loci (51,52). The current SNP data, together with previous expression data, suggest that the DLA region may harbor a genetic risk factor for the development of PAA. Further investigation of the DLA loci is planned. A second region of interest is located on chromosome 7, which has four strongly associated SNPs located within a 1.2 Mb region having 11 genes. None of these genes is known to be involved in autoimmune disorders, although several are expressed in acinar cells.

The data reported here suggest that PAA may be governed by multiple loci with small effects, or it may be a heterogeneous disorder. Although PAA was thought to be a simple autosomal recessive disorder (53,54), the results of a recent test mating between two GSDs with PAA indicate more complex genetic underpinnings (28). The study monitored the cTLI scores for the resulting litter of six over the course of the dogs' lives (8–13 years) and the pancreas for each dog was examined at necropsy. Only two of the six dogs in the litter were diagnosed with PAA (28). Similarly, the clinical presentation of PAA is inconsistent. The age of onset is variable within families, with diagnoses made in

69

puppies and adult dogs alike. The severity of clinical signs also varies, with some dogs responding better to traditional treatment regimens than others. To further complicate the matter, some GSDs never exhibit clinical signs, despite serum cTLI concentrations that are diagnostic for EPI. Variability in age of onset and clinical signs may be an indicator of genetic heterogeneity. Although PAA is the predominant cause of EPI in dogs, chronic pancreatitis, pancreatic cancer, and other underlying medical conditions may also cause EPI.

Variation between processed arrays, a so-called ''batch effect,'' can cause spurious associations (55). The abundance of isolated SNPs detected in our GWAS for PAA is suggestive of such artifacts. However, principal component analysis does not reveal subpopulations that indicate a batch effect. Furthermore, no population stratification is detected between PAA cases and controls (Supplementary Fig. 1). In addition, we included only the robust ''platinum'' SNPs in our analysis to minimize artifacts from incorrect genotype calls.

In this study, we used data from a single cohort of GSDs to identify major loci associated with pituitary dwarfism, DM, and ME. Our data also reveal that the genetics underlying PAA are highly complex. Candidate loci identified herein provide immediate targets for future work.

ACKNOWLEDGEMENTS

REFERENCES

1. Goldbecker WM, Hart EH (1967) This is the German shepherd. T.F.H Publications Inc., Jersey City.

2. Moody JA, Clark LA, Murphy KE (2006) Working dogs: history and applications. In: Ostrander EA, Giger U, Lindblad-Toh K (eds) The dog and its genome. Cold Spring Harbor Laboratory Press, Woodbury, pp 1–18.

3. Wahl JM, Clark LA, Skalli O, Ambrus A, Rees CA et al. (2008) Analysis of gene transcript profiling and immunobiology in Shetland sheepdogs with dermatomyositis. Vet Dermatol 19:52–58.

4. Andresen E, Willeberg P (1976) Pituitary dwarfism in German shepherd dogs: additional evidence of simple, autosomal reces- sive inheritance. Nord Vet Med 28:481–486.

5. Reynaud R, Saveanu A, Barlier A, Enjalbert A, Brue T (2004) Pituitary hormone deficiencies due to transcription factor gene alterations. Growth Horm IGF Res 14:442–448.

6. Lantinga-van Leeuwen IS, Mol JA, Kooistra HS, Rijnberk A, Breen M et al. (2000) Cloning of the canine gene encoding transcription factor Pit-1 and its exclusion as a candidate gene in a canine model of pituitary dwarfism. Mamm Genome 11:31–36.

7. Lantinga-van Leeuwen IS, Kooistra HS, Mol JA, Renier C, Breen M et al. (2000) Cloning, characterization, and physical mapping of the canine Prop-1 gene (PROP1): exclusion as a candidate gene for combined pituitary hormone deficiency in German shepherd dogs. Cytogenet Cell Genet 88:140–144

8. van Oost BA, Versteeg SA, Imholz S, Kooistra HS (2002) Exclusion of the lim homeodomain gene LHX4 as a candidate gene for pituitary dwarfism in German shepherd dogs. Mol Cell Endocrinol 197:57–62.

9. Hanson JM, Mol JA, Leegwater PA, Kooistra HS, Meji BP (2006) The leukemia inhibitory factor receptor gene is not involved in the etiology of pituitary dwarfism in German shepherd dogs. Res Vet Sci 81:316–320.

10. Barclay KB, Haines DM (1994) Immunohistochemical evidence for immunoglobulin and complement deposition in spinal cord lesions in degenerative myelopathy in German shepherd dogs. Can J Vet Res 58:20–24.

11. Averill DR Jr (1973) Degenerative myelopathy in the aging German shepherd dog: clinical and pathologic findings. J Am Vet Med Assoc 162:1045–1051.

12. Coates JR, March PA, Oglesbee M, Ruaux CG, Olby NJ et al. (2007) Clinical characterization of a familial degenerative myelopathy in Pembroke Welsh Corgi dogs. J Vet Intern Med 21:1323–1331.

13. Johnston PE, Barrie JA, McCulloch MC, Anderson TJ, Griffiths IR (2000) Central nervous system pathology in 25 dogs with chronic degenerative radiculomyelopathy. Vet Rec 146:629–633.

14. Awano T, Johnson GS, Wade CM, Katz ML, Johnson GC et al. (2009) Genome-wide association analysis reveals a SOD1 mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis. Proc Natl Acad Sci USA 106: 2794–2799.

15. Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P et al. (1993) Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyo- trophic lateral

sclerosis. Nature 362:59–62.

16. Guilford WG (1990) Megaesophagus in the dog and cat. Semin Vet Med Surg (Small Anim) 5:37–45.

17. Cox VS, Wallace LJ, Anderson VE, Rushmer RA (1980) Hereditary esophageal dysfunction in the miniature Schnauzer dog. Am J Vet Res 41:326–330.

18. Osborne CA, Clifford DH, Jessen C (1967) Hereditary esophageal achalasia in dogs. J Am Vet Med Assoc 151:572–581.

19. Wiberg ME, Saari SA, Westermarck E (1999) Exocrine pancreatic atrophy in German shepherd dogs and rough coated collies: an end result of lymphocytic pancreatitis. Vet Pathol 36:530–541.

20. Westermarck E, Batt RM, Valliant C, Wiberg M (1993) Sequential study of pancreatic structure and function during development of pancreatic acinar atrophy in a German shepherd dog. Am J Vet Res 54:1088–1094.

21. Pfister K, Rossi GL, Freudiger U, Bigler B (1980) Morphological studies in dogs with chronic pancreatic insufficiency. Virchows Arch A 386:91–105.

22. Williams DA, Batt RM (1988) Sensitivity and specificity of radioimmunoassay of serum trypsin-like immunoreactivity for the diagnosis of canine exocrine pancreatic insufficiency. J Am Vet Med Assoc 192:195–201.

23. Räihä M, Westermarck E (1989) The signs of pancreatic degenerative atrophy in dogs and the role of external factors in the ethiology of the disease. Acta Vet Scand 30:447–452.

24. Hall EJ, Bond PM, McLean C, Batt RM, McLean L (1991) A survey of the diagnosis

and treatment of canine exocrine pancreatic insufficiency. J Small Anim Pract 32:613–619.

25. Wiberg ME, Westermarck E (2002) Subclinical exocrine pancreatic insufficiency in dogs. J Am Vet Med Assoc 220:1183–1187.

26. Proschowsky HF, Fredholm M (2007) Exocrine pancreatic insuffi- ciency in the Eurasian dog breed–inheritance and exclusion of two candidate genes. Anim Genet 38:171–173.

27. Wiberg ME (2004) Pancreatic acinar atrophy in German shepherd dogs and rough-coated collies. Etiopathogenesis, diagnosis and treatment. A review Vet Q 26:61–75.

28. Westermarck E, Saari SA, Wiberg ME (2010) Heritability of exocrine pancreatic insufficiency in German shepherd dogs. J Vet Intern Med 24:450–452.

29. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC et al. (2007) Efficient mapping of Mendelian traits in dogs through genome-wide association. Nat Genet 39:1321–1328.

30. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K et al. (2007) A single IGF1 allele is a major determinant of small size in dogs. Science 316:112–115.

31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA et al. (2007) PLINK: a tool set for whole-genome association and population- based linkage analyses. Am J Hum Genet 81:559–575.

32. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300.

33. Schmutz SM, Berryere TG (2007) The genetics of cream coat color in dogs. J Hered

98:544–548.

34. Drögemüller C, Becker D, Brunner A, Haase B, Kircher P et al. (2009) A missense mutation in the SERPINH1 gene in Dachshunds with osteogenesis imperfecta. PLoS Genet 5:e1000579.

35. Gill JL, Tsai KL, Krey C, Noorai RE, Vanbellinghen JF et al. (2011) A canine BCAN microdeletion associated with episodic falling syndrome. Neurobiol Dis. 45:130-136.

36. Bach I, Rhodes SJ, Pearse RV II, Heinzel T, Gloss B et al. (1995) P-Lim, a LIM homeodo- main factor, is expressed during pituitary organ and cell commitment and synergizes with Pit-1. Proc Natl Acad Sci USA 92:2720–2724.

37. Netchine I, Sobrier ML, Krude H, Schnabel D, Maghnie M et al. (2000) Mutations in LHX3 result in a new syndrome revealed by combined pituitary hormone deficiency. Nat Genet 25:182–186.

38. Pfaeffle RW, Savage JJ, Hunter CS, Palme C, Ahlmann M et al. (2007) Four novel mutations of the LHX3 gene cause combined pituitary hormone deficiencies with or without limited neck rotation. J Clin Endocrinol Metab 92: 1909–1919.

39. Bhangoo APS, Hunter CS, Savage JJ, Anhalt H, Pavlakis S et al. (2006) Clinical case seminar: a novel LHX3 mutation presenting as combined pituitary hormonal deficiency. J Clin Endocr Metab 91:747–753.

40. Rajab A, Kelberman D, de Castro SC, Biebermann H, Shaikh H et al. (2008) Novel mutations in LHX3 are associated with hypopituitarism and sensorineural hearing loss. Hum Mol Genet 17:2150–2159.

41. Kooistra HS, Voorhout G, Mol JA, Rijnberk A (2000) Combined pituitary hormone

deficiency in German shepherd dogs with dwarfism. Domest Anim Endocrin 19:177–190.

42. Sheng HZ, Zhadanov AB, Mosinger B Jr, Fujii T, Bertuzzi S et al. (1996) Specification of pituitary cell lineages by the LIM homeobox gene LHX3. Science 272:1004–1007.

43. Lane PW, Dickie MM (1968) Three recessive mutations producing disproportionate dwarfing in mice: achondroplasia, brachymor- phic, and stubby. J Hered 59:300–308.

44. Colvin SC, Malik RE, Showalter AD, Sloop KW, Rhodes SJ (2011) Model of pediatric pituitary hormone deficiency separates the endocrine and neural functions of the LHX3 transcription factor in vivo. Proc Natl Acad Sci USA 108:173–178.

45. Voorbij AM, Leegwater PA, Kooistra HS (2010) Hypopituitarism associated dwarfism in German shepherds, saarloos wolf dogs and Czechoslovakian wolf dogs. Access to genetic testing. Tijdschr Diergeneesk 135:950–954.

46. Clark LA, Wahl JM, Steiner JM, Zhou W, Ji W et al. (2005) Linkage analysis and gene expression profile of pancreatic acinar atrophy in the German shepherd dog. Mamm Genome 16:955–962.

47. Kennedy LJ, Davison LJ, Barnes A, Short AD, Fretwell N et al. (2006) Identification of susceptibility and protective major histocompatibility complex haplotypes in canine diabetes mellitus. Tissue Antigens 68: 467–476.

48. Barnes A, O'Neill T, Kennedy LJ, Short AD, Catchpole B et al. (2009) Association of canine anal furunculosis with TNFA is secondary to linkage disequilibrium with DLA-DRB1*. Tissue Antigens 73:218–224.

49. Greer KA, Wong AK, Liu H, Famula TR, Pedersen NC et al. (2010) Necrotizing meningoencephalitis of Pug dogs associates with dog leukocyte antigen class II and resembles acute variant forms of multiple sclerosis. Tissue Antigens 76:110–118.

50. Hughes AM, Jokinen P, Bannasch DL, Lohi H, Oberbauer AM (2010) Association of a dog leukocyte antigen class II haplotype with hypoadrenocorticism in Nova Scotia duck tolling retrievers. Tissue Antigens 75:684–690.

51. Wilbe M, Jokinen P, Hermanrud C, Kennedy LJ, Strandberg E et al. (2009) MHC class II polymorphism is associated with a canine SLE-related disease complex. Immunogenetics 61:557–564.

52. Wilbe M, Jokinen P, Truve ́K, Seppala EH, Karlsson EK et al. (2010) Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. Nat Genet 42:250–254.

53. Moeller EM, Steiner JM, Clark LA, Murphy KE, Famula TR et al. (2002) Inheritance of pancreatic acinar atrophy in German shepherd dogs. Am J Vet Res 63(10):1429–1434.

54. Westermarck E (1980) The hereditary nature of canine pancreatic degenerative atrophy in the German shepherd dog. Acta Vet Scand 21:389–394.

55. Hong H, Su Z, Ge W, Shi L, PerkinsR et al. (2008) Assessing batch effects of genotype calling algorithm BRLMM for the Affymetrix Gene- Chip human mapping 500 K array set using 270 HapMap samples. BMC Bioinformatics 9:S17.

CHAPTER IV

GENOME-WIDE ASSOCIATION MAPPING AND IDENTIFICATION OF
CANDIDATE GENES FOR THE RUMPLESS AND EAR-TUFTED
TRAITS OF THE ARAUCANA CHICKEN

Rooksana E. Noorai[a], Nowlan H. Freese[b], Lindsay M. Wright[a], Susan C. Chapman[b],

Leigh Anne Clark[a]

[a] Department of Genetics and Biochemistry, College of Agriculture, Forestry and Life
Sciences, Clemson University, Clemson, SC 29634, USA

[b] Department of Biological Sciences, College of Agriculture, Forestry and Life Sciences,
Clemson University, Clemson, SC 29634, USA

ABSTRACT

Araucana chickens are known for their rounded, tailless rumps and tufted ears. Inheritance studies have shown that the rumpless (*Rp*) and ear-tufted (*Et*) loci each act in an autosomal dominant fashion, segregate independently, and are associated with an increased rate of embryonic mortality. To find genomic regions associated with *Rp* and *Et*, we generated genome-wide SNP profiles for a diverse population of 60 Araucana chickens using the 60K chicken SNP BeadChip. Genome-wide association studies using 40 rumpless and 11 tailed birds showed a strong association with rumpless on Gga 2 (P_{raw} = 2.45 x 10^{-10}, P_{genome} = 0.00575), and analysis of genotypes revealed a 2.14 Mb haplotype shared by all rumpless birds. Within this haplotype, a 0.74 Mb critical interval containing two *Iroquois* homeobox genes, *Irx1* and *Irx2*, was unique to rumpless Araucana chickens. *Irx1* and *Irx2* are central for developmental prepatterning, but neither gene is known to have a role in mechanisms leading to caudal development. A second genome-wide association analysis using 30 ear-tufted and 28 non-tufted birds revealed an association with tufted on Gga 15 (P_{raw} = 6.61 x 10^{-7}, P_{genome} = 0.0981). We identified a 0.58 Mb haplotype common to tufted birds and harboring 7 genes. Because homozygosity for *Et* is nearly 100% lethal, we employed a heterozygosity mapping approach to prioritize candidate gene selection. A 60 kb region heterozygous in all Araucana chickens contains the complete coding sequence for *TBX1* and partial sequence for *GNB1L*. *TBX1* is an important transcriptional regulator of embryonic development and a key genetic determinant of human DiGeorge syndrome. Herein, we describe localization of *Rp* and *Et* and identification of positional candidate genes.

80

INTRODUCTION

There are hundreds of domestic chicken breeds worldwide (1). Breeds were generally developed for meat and egg production, but morphological traits, plumage color, and other distinctive characteristics were also selected. The Araucana chicken, originally from Chile, is a multi-purpose breed initially established for its blue-shelled eggs (1,2). Araucana chickens are also known for two other distinguishing traits: a rounded, tailless rump and protruding ear-tufts. Although these traits segregate in the population, the United States Araucana breed standard requires show birds to possess both phenotypes.

The rumpless phenotype is characterized by the absence of all free caudal vertebrae and the uropygial gland (3). Without underlying skeletal support, birds with caudal truncation lack a fleshy rump and tail feathers (3). An intermediate rumpless phenotype, wherein some caudal vertebrae are present but irregularly fused together, is thought to result from a modifier gene introduced through crosses with non-Araucana tailed chickens (3,4). The rumpless phenotype arises from a defect in caudal patterning that is controlled by a dominant gene (Rp) (3). Rumpless Araucana chickens may be heterozygous or homozygous for this locus. In test matings, all rumpless intermediates were determined to be heterozygous (Rp/rp^+) (3). Homozygosity is underrepresented among chicks from rumpless to rumpless matings, indicating that the Rp/Rp genotype has reduced viability (3,5). Birds having at least one copy of Rp have increased mortality in

81

the embryo stage, with death occurring at 17 to 21 days of incubation (3). Rumpless birds also have reduced fecundity as adults (3).

Ear-tufts are feather-covered, epidermal protrusions originating near the ear canal (Fig. 1). The mass of tissue forming the protrusion, or peduncle, is believed to develop as a result of the incomplete fusion of the hyomandibular arches, and it can vary in position and length (from two mm to two cm) (6,7). Tufted chickens may also have structural rearrangement of the ears (6). Abnormalities include irregularly shaped external ear openings and shortened or absent external auditory canals (6).
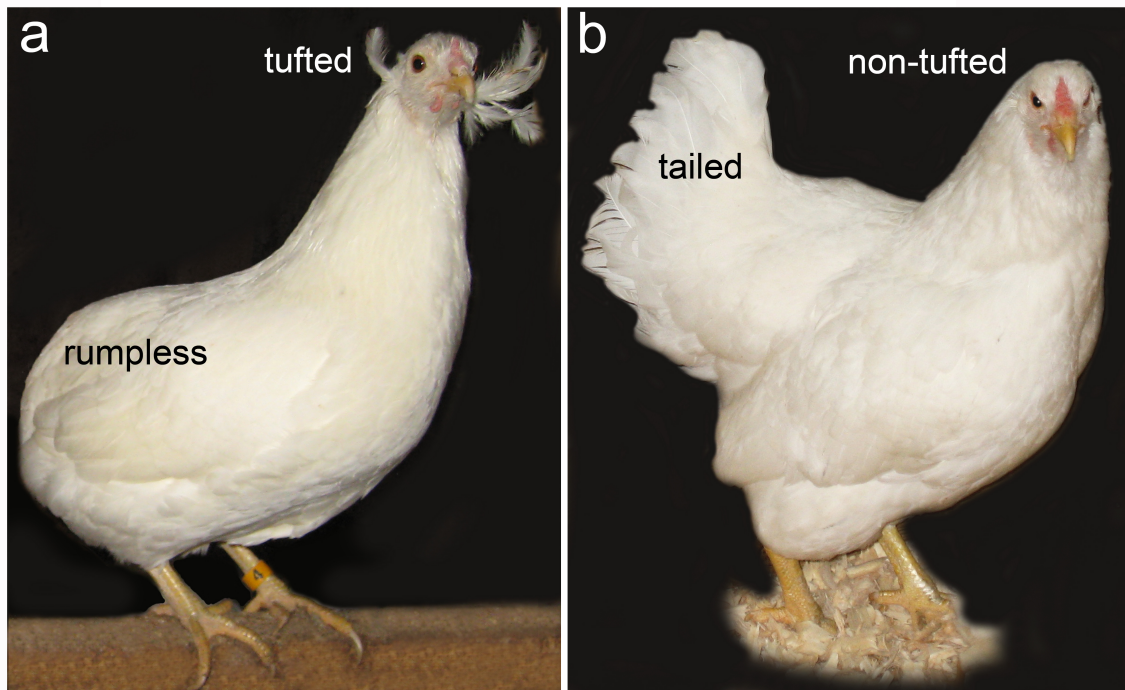
Fig. 1. Araucana chicken. (a) General appearance of a rumpless, tufted Araucana chicken. (b) For comparison, a tailed, non-tufted Araucana chicken.
doi:10.1371/journal.pone.0040974.g001

Inheritance studies indicate that tufted is governed by a dominant locus, Et (6,8). Test matings show that all tufted birds are heterozygous (Et/et^+) and that homozygosity for Et is lethal at about 17-19 days of incubation (6,8). Lethality among a portion of heterozygous birds is also reported, appearing to occur at 20-21 days of incubation (8). Post-hatch mortality is significantly higher among tufted chickens (6,8).

Because tufts can occur unilaterally or bilaterally and may differ in size from one side to the other, Et is proposed to have variable expressivity (6). In addition, a paucity of tufted progeny from mating studies in 1978 suggests reduced penetrance of the tufted locus (6). In 1981, Somes and Pabilonia identified a tufted male that produced excessive tufted progeny when crossed with an et^+/et^+ White Leghorn (86%), and they speculated that Et/Et birds may occasionally reach maturity (8). The non-tufted chicks from the Et/Et male produced tufted progeny when crossed with an et^+/et^+ White Leghorn, indicating that their predicted genotype does not match their phenotype, providing further evidence for variable penetrance.

The aim of our investigation was to localize the genetic bases for the rumpless and tufted phenotypes of the Araucana chicken. To this end, we generated genome-wide SNP profiles for 60 Araucana chickens using the 60K chicken SNP BeadChip (9). Using a genome-wide association approach, we elucidate the chromosomal regions harboring Rp and Et and identify strong candidate genes for each trait.

RESULTS

Case/control analyses were carried out using 40 rumpless and 11 tailed Araucana chickens (Fig. 2a). Seven birds described as having partial tails by their breeders were
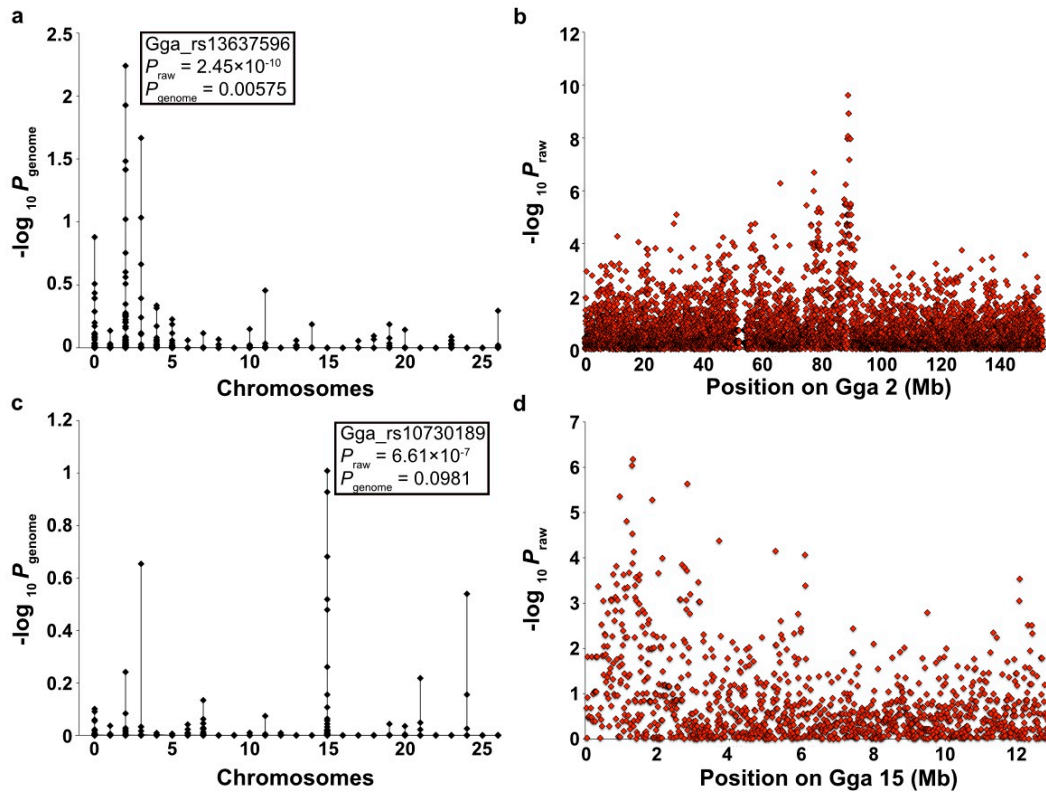
Fig. 2. Genome-wide association for Rp and Et. After 100,000 permutations, the genome-wide adjusted P values (2log10 Pgenome) for each SNP are plotted by chromosome (left). The raw P values for the most strongly associated chromosomes are plotted against chromosomal position (right). (a,b) 40 rumpless versus 11 tailed Araucana chickens (c,d) 30 tufted versus 28 non-tufted Araucana chickens. doi:10.1371/journal.pone.0040974.g002

85

excluded from the rumpless association analysis because of uncertainty concerning their phenotype. A total of 191 SNPs were associated with the rumpless phenotype ($P_{raw} \leq$ 0.0001), 72 of which were located on Gga 2 (Fig. 2b). The most significant result obtained was for SNP Gga_rs13637596, located on chromosome 2 at position 88.95 Mb (P_{raw} = 2.45 x 10^{-10}, P_{genome} = 0.00575). The next two most significant results were for proximal SNPs located at 89.17 Mb (P_{raw} = 1.20 x 10^{-9}, P_{genome} = 0.0119) and 89.19 Mb (P_{raw} = 1.20 x 10^{-9}, P_{genome} = 0.0119).

Analysis of genotypes in the Gga 2 region revealed a 2.14 Mb haplotype (87.99 – 90.13 Mb) predicted to contain five genes (Fig. 3). All 40 rumpless birds had at least one copy of the haplotype: 18 were homozygous and 22 were heterozygous. Partial tailed birds were heterozygous. The haplotype was absent in its entirety from the 11 tailed birds. Three tailed birds were heterozygous for partial blocks of the haplotype and further delimit the critical interval to 0.74 Mb (88.77 - 89.51 Mb). This region contains two candidate genes: *Irx1* and *Irx2*.
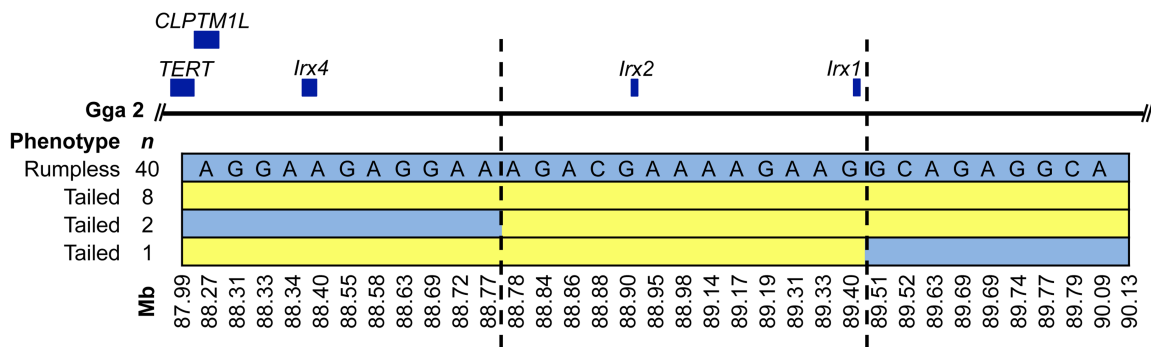
Fig. 3. Localization of Rp. Physical map showing the relative positions of mapped genes and informative SNP markers within the 2.14 Mb rumpless haplotype on Gga 2. Light blue shading denotes the rumpless haplotype (alleles are shown in the top row). Dashed lines flank the critical interval wherein no tailed birds share the rumpless haplotype.
doi:10.1371/journal.pone.0040974.g003

Analyses for association with the tufted phenotype, using 30 cases and 28 controls, resulted in 31 significant SNPs, 11 of which map to Gga 15 (Fig. 2c). The most significant results were for SNPs Gga_rs10730189 (P_{raw} = 6.61 x 10^{-7}, P_{genome} = 0.0981) and Gga_rs15762547 (P_{raw} = 9.19 x 10^{-7}, P_{genome} = 0.118), located at positions 1.33 Mb and 1.30 Mb on chromosome 15, respectively. Four other proximal SNPs also reached significance (Fig. 2d).

Analysis of genotypes reveals that 29 of 30 tufted birds shared a haplotype extending from the telomere of Gga 15 to position 1.75 Mb. These birds were heterozygous for the complete haplotype. Two of 28 non-tufted birds were also heterozygous for the haplotype in its entirety. A single tufted bird shared only part of the 1.75 Mb haplotype, defining a 0.58 Mb (0.90 – 1.48 Mb) critical interval that is

87

heterozygous in all 30 tufted birds and contains 7 genes. Because tufted is nearly always recessive lethal, blocks of homozygosity for the tufted haplotype were identified to reduce the number of candidate genes. Homozygosity blocks in three birds flank a 60 kb interval harboring two genes: *TBX1* and *GNB1L* (Fig. 4).
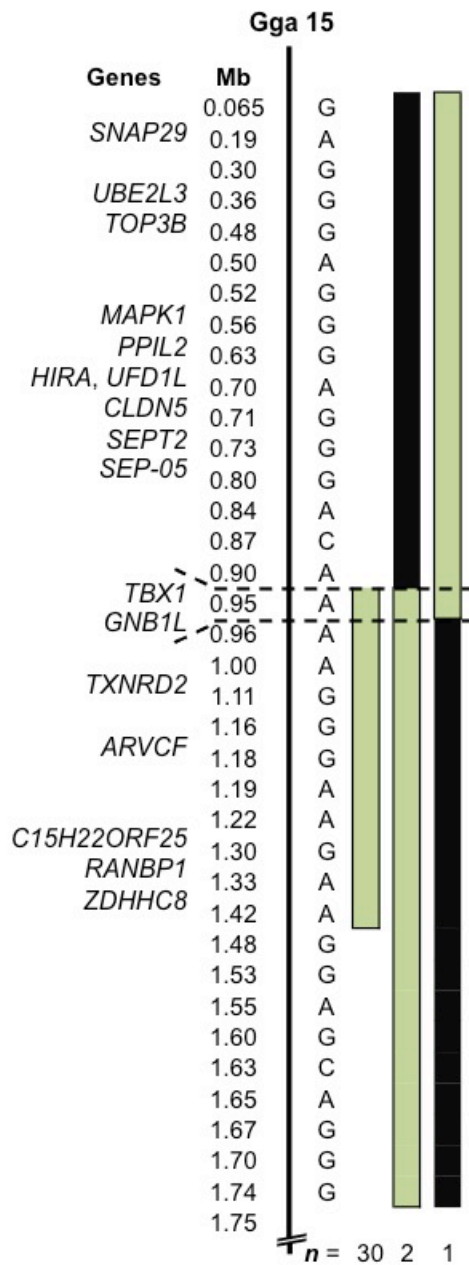
Fig. 4. Localization of Et. Physical map showing the relative positions of genes and informative SNP markers in the associated region of Gga 15. Alleles of the tufted haplotype and positions are shown. Pale green bars denote heterozygosity for the tufted haplotype. Black bars denote homozygosity for the tufted haplotype. Dashed lines mark a 60 kb interval wherein all tufted birds are heterozygous for the haplotype.
doi:10.1371/journal.pone.0040974.g004

DISCUSSION

In this study, we used genome-wide SNP profiles to localize genes causative for two breed-defining phenotypes of Araucana chickens, rumpless and ear-tufts. We took advantage of the fact that both traits segregate independently in the population by using a single data set to carry out an association analysis for each trait. Haplotype analyses based on inheritance patterns were used to identify positional candidate genes for both traits.

We identified a rumpless haplotype spanning 2.14 Mb and five genes on chromosome 2. The haplotype is present in the heterozygous or homozygous state in rumpless birds. All 7 birds with partial tails are heterozygous for the rumpless haplotype and likely represent the intermediate phenotype described by Dunn and Landauer (3). Because rumpless is dominant and fully penetrant, we further delimited the critical interval by identifying regions of the haplotype shared by tailed birds. A 0.74 Mb region common to all rumpless birds, and absent from 11 tailed birds, harbors *Rp*.

These data reveal that *Rp* maps to a region of Gga 2 that is distinct from the predicted location of genes previously associated with caudal truncation (10-14). The 0.74 Mb critical interval contains the *Iroquois* homeobox genes, *Irx1* and *Irx2*. The *Iroquois* genes encode transcription factors that function in patterning and regionalization of tissues early in development (15). *Irx1* and *Irx2* are prepattern and proneural genes first identified in *Drosophila* and *Xenopus* (16,17). Studies of gene function suggest that

90

Irx genes have redundant yet distinct roles in development (18,19). *Irx* genes have been knocked out in mice and zebrafish with little effect on tail development (19-23). However, the rumpless phenotype is dominant, suggesting that misexpression of *Irx1* or *Irx2* may underlie the trait, rather than loss of function.

We identified SNPs on Gga 15 that are strongly associated with the tufted phenotype and define a 0.58 Mb haplotype for which all tufted birds in our cohort are heterozygous. No birds are homozygous for the complete tufted haplotype. These data support conclusions from previous inheritance studies that suggest nearly 100% of tufted birds are heterozygous, and that *Et/Et* is lethal (6,8).

Two non-tufted Araucana chickens are heterozygous for the tufted haplotype. These birds may signify reduced penetrance. Penetrance of the tufted allele is estimated to range from 86% to 96% (6,8). Based on the assigned phenotypes and the associated haplotype, we observed 94% penetrance in our cohort. Alternatively, these birds may have been incorrectly phenotyped by their breeders due to short peduncles or missing protruding feathers.

The 0.58 Mb haplotype harbors 7 protein-coding genes. Unlike rumpless, identification of the tufted haplotype in non-tufted birds could not be used to narrow the critical interval because of reduced penetrance. However, because homozygosity for *Et* is nearly always lethal, we were able to prioritize candidate gene selection using

heterozygosity mapping. Tufted birds with blocks of homozygosity extending into the 0.58 Mb common haplotype were identified, and these regions were deemed less likely to harbor the *Et* locus. These data indicate that *Et* is located in a region containing partial coding sequence for *GNB1L*, which encodes a protein implicated in neuropsychiatric disorders (24,25), and complete coding sequence for *TBX1* (26), an important transcriptional regulator of embryonic development.

Haploinsufficiency for *TBX1* is considered to be the key genetic determinant of human DiGeorge syndrome (DGS), which is caused by a heterozygous chromosomal deletion of 22q11.2 (27). While the clinical phenotype is highly variable, DGS is characterized by craniofacial and cardiovascular abnormalities. Malformations in DGS are attributed to disturbed segmentation and patterning of the pharyngeal structures (28). Auricular defects common in DGS include narrow or absent external ear canal and protruding ears (29). Homozygosity for null mutations of *TBX1* in mice and zebrafish causes a range of phenotypic effects similar to DGS, including abnormal ear development (30,31). Based on phenotypic similarities between the malformations causing ear tufts and DGS, *TBX1* is a highly plausible candidate gene and the primary focus of ongoing work to identify the genetic basis for ear-tufts in Araucana chickens.

In conclusion, we used genome-wide association and haplotype analyses to localize *Rp* and *Et* to chicken chromosomes 2 and 15, respectively. In addition, we identified candidate genes that are immediate targets for future work.

MATERIALS AND METHODS

Ethics Statement

This study was approved by the Clemson University IACUC protocol number 2011-041 and IBC protocol number 2010-041.

Study Cohort

Whole blood for DNA was collected from 6 different flocks of Araucana chickens from the United States. Phenotypic information and photographs, when available, were provided by owners. Birds with tufts of any size and on either side of the head were classified as tufted. Because both traits segregate in the Araucana population, birds were selected to ensure that the phenotypes were balanced. Our study cohort comprised 60 Araucana chickens: 21 rumpless/tufted birds, 20 rumpless/non-tufted birds, 7 tailed/non-tufted birds, five tailed/tufted birds, five partial/tufted birds, and two partial/non-tufted birds. Genomic DNA was isolated using the DNeasy blood and tissue kit (QIAGEN, Valencia, USA) and adjusted to a concentration of 50 ng/μL.

Genome-wide Association Mapping

SNP genotypes were generated using the Illumina 60K chicken SNP BeadChip, which has 57,636 SNPs across chromosomes 1 through 28, Z, W, and two unmapped linkage groups (9). BeadChips were processed by DNA Landmarks (Quebec, Canada), according to manufacturer's protocols. Raw data files were analyzed using GenomeStudio's Genotyping Module to generate SNP calls. The PLINK Input Report

Plug-in v2.1.1 was used to format the data. For analysis, Gga 27, Gga 28, Gga Z, Gga W, and microchromosomes were all identified as chromosome zero. Case/control analyses using 56,685 SNPs were performed using PLINK (32). Two birds with excessive missing data were excluded from all analyses. By convention, P_{raw} values ≤ 0.0001 were considered significant. Permutation testing, using 100,000 iterations, was carried out using PLINK.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: NHF SCC LAC. Performed the experiments: REN NHF. Analyzed the data: REN NHF LMW LAC. Contributed reagents/materials/analysis tools: SCC LAC. Wrote the paper: REN SCC LAC.

REFERENCES

1. Ekarius C (2007) Storey's Illustrated Guide to Poultry Breeds. China: Storey Publishing. pp. 23-24.

2. Browman DL (1978) Advances in Andean Archaeology. Great Britain: Mouton Publishers. pp. 189-196.

3. Dunn LC, Landauer W (1934) The genetics of the rumpless fowl with evidence of a case of changing dominance. J Genet 29: 217-243.

4. Dunn LC, Landauer W (1936) Further data on genetic modification of rumplessness in the fowl. J Genet 33: 401-405.

5. Zwilling E (1942) The development of dominant rumplessness in chick embryos. Genetics 27: 641-656.

6. Somes Jr RG (1978) Ear-Tufts: a skin structure mutation of the Araucana fowl. J Hered 69: 91-96.

7. Pabilonia MS, Somes Jr RG (1983) The Embryonic Development of Ear-Tufts and Associated Structural Head and Neck Abnormalities of the Araucana Fowl. Poult Sci 62: 1539-1542.

8. Somes Jr RG, Pabilonia MS (1981) Ear tuftedness: a lethal condition in the Araucana fowl. J Hered 72: 121-124.

9. Groenen MAM, Megens HJ, Zare Y, Warren WC, Hillier LW et al. (2011) The development and characterization of a 60K SNP chip for chicken. BMC Genomics 12: 274.

10. Herrmann BG, Labeit S, Poustka A, King TR, Lehrach H (1990) Cloning of the T gene required in mesoderm formation in the mouse. Nature 343: 617-622.

11. Greco TL, Takada S, Newhouse MM, McMahon JA, McMahon AP et al. (1996) Analysis of the vestigial tail mutation demonstrates that Wnt-3a gene dosage regulates mouse axial development. Genes Dev 10: 313-324.

12. Ross AJ, Ruiz-Perez V, Wang Y, Hagan DM, Scherer S et al. (1998) A Homeobox gene, HLXB9, is the major locus for dominantly inherited sacral agenesis. Nat Genet 20: 358-361.

13. Abu-Abed S, Dollé P, Metzger D, Beckett B, Chambon P et al. (2001) The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. Genes Dev 15: 226-240.

14. van den Akker E, Forlani S, Chawengsaksophak K, de Graaff W, Beck F et al. (2002) *Cdx1* and *Cdx2* have overlapping functions in anteroposterior patterning and posterior axis elongation. Development 129: 2181-2193.

15. Cavodeassi F, Modolell J, Gómez-Skarmeta JL (2001) The Iroquois family of genes: from body building to neural patterning. Development 128: 2847-2855.

16. Gómez-Skarmeta JL, Diez del Corral R, de la Calle-Mustienes E, Ferré-Marcó D, Modolell J (1996) Araucan and caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. Cell 85: 95-105.

17. Gómez-Skarmeta JL, Modolell J (1996) araucan and caupolican provide a link between compartment subdivisions and patterning of sensory organs and veins in the Drosophila wing. Genes Dev 10: 2935-2945.

18. Costantini DL, Arruda EP, Agarwal P, Kim KH, Zhu Y et al. (2005) The homeodomain transcription factor Irx5 establishes the mouse cardiac ventricular repolarization gradient. Cell 123: 347-358.

19. Lebel M, Agarwal P, Cheng CW, Kabir MG, Chan, TY (2003) The Iroquois homeobox gene Irx2 is not essential for normal development of the heart and midbrain-hindbrain boundary in mice. Mol Cell Biol 23: 8216-8225.

20. Itoh M, Kudoh T, Dedekian M, Kim CH, Chitnis AB (2002) A role for *iro1* and *iro7* in the establishment of an anteroposterior compartment of the ectoderm adjacent to the midbrain-hindbrain boundary. Development 129: 2317-2327.

21. Peters T, Ausmeier K, Dildrop R, Rüther U (2002) The mouse Fused toes (Ft) mutation is the result of a 1.6-Mb deletion including the entire Iroquois B gene cluster. Mamm Genome 13: 186-188.

22. Cheng CW, Yan CHM, Hui CC, Strähle U, Cheng SH (2006) The Homeobox gene *irx1a* is required for the propagation of the neurogenic waves in the zebrafish retina. Mech Develop 123: 252-263.

23. Kimura W, Machii M, Xue X, Sultana N, Hikosaka K et al. (2011) Irx1 mutant mice show reduced tendon differentiation and no patterning defects in musculoskeletal system development. Genesis 49: 2-9.

24. Williams NM, Glaser B, Norton N, Williams H, Pierce T et al. (2008) Strong evidence that *GNB1L* is associated with schizophrenia. Hum Mol Genet 17: 555-566.

25. Li Y, Zhao Q, Wang T, Liu J, Li J et al. (2011) Association study between *GNB1L* and three major mental disorders in Chinese Han populations. Psychiat Res 187: 457-459.

26. Völker M, Backström N, Skinner BM, Langley EJ, Bunzey SK et al. (2010) Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. Gen Res 20: 503-511

27. Yagi H, Furutani Y, Hamada H, Sasaki T, Asakawa S, et al. (2003) Role of *TBX1* in human del22q11.2 syndrome. Lancet 362: 1366-1373.

28. Wurdak H, Ittner LM, Sommer L (2006) DiGeorge syndrome and pharyngeal apparatus development. BioEssays 28: 1078-1086.

29. Butts SC (2009) The facial phenotype of the velo-cardio-facial syndrome. Int J Pediatr Otorhinolaryngol 73: 343-350.

30. Jerome LA, Papaioannou VE (2001) DiGeorge syndrome phenotype in mice mutant for the T-box gene, *Tbx1*. Nat Genet 27: 286-291.

31. Piotrowski T, Ahn DG, Schilling TF, Nair S, Ruvinsky I et al. (2003) The zebrafish *van gogh* mutation disrupts *tbx1*, which is involved in the DiGeorge deletion syndrome in humans. Development 130: 5043-5052.

32. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575.

CHAPTER V

GENOME-WIDE ASSOCIATION STUDY IN COLLIES IDENTIFIES A NOVEL LOCUS FOR DERMATOMYOSITIS

Rooksana E. Noorai, Leigh Anne Clark

Department of Genetics and Biochemistry, College of Agriculture, Forestry and Life Sciences, Clemson University, Clemson, SC 29634, USA

Unpublished

ABSTRACT

Canine familial dermatomyositis (DM) is a spontaneous inflammatory myopathy that predominantly affects the collie and Shetland sheepdog breeds. DM is characterized by painful lesions on the face and extremities; and, is similar clinically, histologically, and immunologically to human juvenile dermatomyositis (human DM). The population structure of dogs makes them an excellent model for the study of complex hereditary diseases and DM in collie and Shetland sheepdogs is the only animal model described to date. To find genomic regions associated with DM, we generated genome-wide SNP profiles for a diverse population of 47 collies using the Illumina Canine HD BeadChip. Genome-wide association studies using 17 DM-affected and 14 healthy control collies showed a single, strong association with DM on CFA 10 ($P_{raw} = 1.97 \times 10^{-6}$), and analysis of genotypes revealed a 10.5 Mb haplotype shared by all DM-affected collies. Currently, there are only five mapped genes in this region of the canine reference genome; however, based on conservation of synteny with human chromosome 12, this region could contain as many as 130 genes. Herein, we describe the localization of DM and discuss positional candidate genes having a role in immune function.

INTRODUCTION

Juvenile dermatomyositis (human DM) is the most common childhood inflammatory myopathy, comprising about 85% of all diagnosed cases of inflammatory myopathy (1,2). Human DM, which affects 3.2 children per million between the ages of 2 and 17 in the United States (with similar numbers in the United Kingdom), is a vasculopathy that causes a characteristic skin rash and debilitating muscle weakness (3-5). The current method of diagnosis is based on a criterion and classification system that was proposed in 1975 (6,7). In diagnosing human DM, the Bohan and Peter system considers five major criteria, which are muscle involvement in the trunk with or without respiratory involvement, signs of severe inflammation in muscle biopsy, elevated levels of muscle enzymes in the blood, myopathy diagnosed via electrocardiogram, and characteristic dermatologic involvement. Definitive diagnosis requires the presence of three criteria in addition to dermatologic signs; probable diagnosis requires two criteria in addition to dermatologic signs, while possible diagnosis requires one criterion in addition to dermatologic signs (6,7). Early diagnosis and proper treatment of the disease are important for improving the prognosis for an affected child (8). Calcinosis, abnormal calcium deposits in various tissues, has been seen in patients with a long history of undiagnosed disease, inadequate therapy, or chronic disease (9,10). Unfortunately, even with well-defined criteria, a diagnosis of human DM can be hindered by the variety of symptoms of a patient's initial presentation (8). Once diagnosed, treatment with corticosteroids and/or immunosuppressive drugs is necessary for remission or to prevent further deterioration (8,9,11-13). The prognosis for a child with human DM is positively

101

correlated with a quick diagnosis and proper treatment (9). The impact of human DM on the life of a patient varies, but many children continue to have chronic flare-ups and/or suffer from the chronic rashes, lesions, and muscle weakness (8). The etiology of DM is not known, but DM is thought to be an autoimmune disease brought on by infection with a virus or other agent (14-17). In one study, 71% of human DM patients had a clinical history consistent with an infection prior to initial signs of human DM (18). There may be a gene or multiple genes that predispose a patient to the condition (19).

Canine familial dermatomyositis (DM) is a naturally occurring autoimmune disease of domestic dogs that is similar in clinical presentation to human DM. DM is characterized by an inflammatory response in the skin and muscle tissue, typically around the face, paws, and tail. The disease phenotype is highly variable, with signs ranging from a mild, transient presentation to a near fatal presentation. Clinical findings consistent with DM include the development of skin lesions such as hair loss, redness, scaling and crusting on the face, ears, legs, and tail tip. Dogs afflicted with DM often develop disfiguring scars and secondary skin infections. In severe cases, inflammation affects muscle tissue and can cause muscle atrophy, megaesophagus, aspiration pneumonia, and difficulty eating or walking. Muscle involvement can cause a poor quality of life and owners often elect for euthanasia. Unfortunately, there is no cure for DM and clinical signs may reoccur. Onset of DM ranges from two months to several years of age. A dog may have produced multiple litters before the onset of clinical signs, making it difficult for breeders to control disease transmission. Current treatment for DM

includes the management of clinical signs, such as severe pain and secondary skin infections. A twice-daily dose of the drug Pentoxifylline, which increases blood circulation, is also recommended (20).

The diagnosis of DM is made by a veterinary histopathologist using skin punch biopsies of actively afflicted tissue. All three layers of skin (dermis, epidermis, and subcutis) are evaluated for signs of inflammation (Dr. J. Mansell, personal communication). Some characteristic signs of DM include vacuolar alteration and dilation of the basal cell layer in the epidermis, signs of apoptosis, and sparse superficial perivascular lymphocytic infiltrates. When these and other cell changes are observed, then a diagnosis of DM is made (Dr. J. Mansell, personal communication).

Two breeds with the highest prevalence of DM are collies and Shetland sheepdogs (Shelties). DM has been diagnosed and characterized in both breeds since 1983 (21-23). These two herding breeds share a recent common ancestor (24) and likely share the same original mutation contributing to DM. The exact prevalence of DM is unknown, but affected dogs have been found in different breeding lines across the United States. Data from two studies suggest that the inheritance pattern of DM is autosomal dominant (21, Sherry Lindsey, personal communication). Hargis and colleagues, (21) outcrossed an affected male purebred collie and an unaffected female purebred Labrador retriever. Three of the four offspring developed clinical signs of DM that were milder than those seen in offspring from an inbred litter. The inbred litter was a cross of the

103

aforementioned collie sire and an affected female purebred collie. In an unpublished study, two DM-affected purebred Shelties were bred and produced a single puppy that never developed clinical signs of DM (Sherry Lindsey, personal communication).

Genome-wide association studies (GWAS) have been successfully used to identify simple and complex traits of the dog (25,26). The goal of this work is to identify a genetic region(s) of association with DM, utilizing GWAS.

MATERIALS AND METHODS

Sample Collection

Local and national collie breeders and collie rescues were contacted by electronic methods to solicit research subjects. Whole blood or buccal cells were collected from participants by their regular veterinarians. DM-affected collies were required to have a histopathology report from a skin punch biopsy with a positive or differential diagnosis of DM. Pedigrees from DM affected collies were collected when available. Control collies were required to be at least seven years of age or older because, while variable, the onset of DM usually occurs prior to 6 years of age. Participants had to be pedigreed, without history of a dermatologic condition, and unrelated to any DM affected or control collies within three generations.

Genome-wide Association Mapping

Genomic DNA was isolated using the DNeasy blood and tissue kit (QIAGEN, Valencia, USA) and adjusted to a concentration of 100 ng/μL. Illumina CanineHD Infinium BeadChips were used to profile 173,662 SNP genotypes per collie, across autosomal chromosomes 1 through 38 and sex chromosomes 39/X and 40/Y. BeadChips were processed at the Cornell University Life Sciences Core Laboratories Center (Ithaca, New York, USA) according to manufacturer's protocols. Raw data files were analyzed using GenomeStudio's Genotyping Module to generate SNP calls. The PLINK Input Report Plug-in v2.1.1 was used to format the data. Case/control analyses were conducted with PLINK (27). By convention, P_{raw} values ≤ 0.0001 were considered significant. R (28) was used to generate Manhattan plots from the $-\log_{10}$ of the P_{raw} value for each SNP. Permutation testing, using 100,000 iterations, was carried out using PLINK. BeadChip SNP positions were remapped to canFam3.

RESULTS

Our study cohort consisted of 27 DM-affected (ten females/17 males) and 20 (12 females/eight males) healthy collies. The age of DM-affected collies ranged from three months to 11 years (at the time of sampling), while control dogs were seven to 12 years of age. All blood and buccal samples were obtained from the United States, except for a single DM-affected collie residing in Australia (no American ancestors in X # generations). There was no significant difference in the distribution of males and females in this study (p= 0.1476). We were unable to obtain information regarding the severity of

the clinical signs of DM in all cases.

A principal components analysis using all 173,662 SNPs identified five collies that did not cluster with the majority of the population, indicating a level of population stratification from the remaining collies (Fig. 1). These five were considered outliers, one control and four DM-affected collies, and were excluded from further analyses. The remaining 42 dogs showed no evidence for population stratification.

To demonstrate the utility of the assembled cohort, a positive control was mapped using GWAS. Hair length in dogs is determined by mutations in *FGF5* on chromosome 32 (29). Smooth in collies refers to a short hair length, which is dominant to rough (long) hair. Case/control analyses for eight smooth (six DM-affected/two controls) vs 33 rough (15 DM-affected/18 controls) collies were carried out (Fig. 2). The most significantly associated SNP was BICF2G630600757 on chromosome 32 ($P_{raw} = 4.70 \times 10^{-12}$), and it is located approximately 180 kb downstream from *FGF5*.
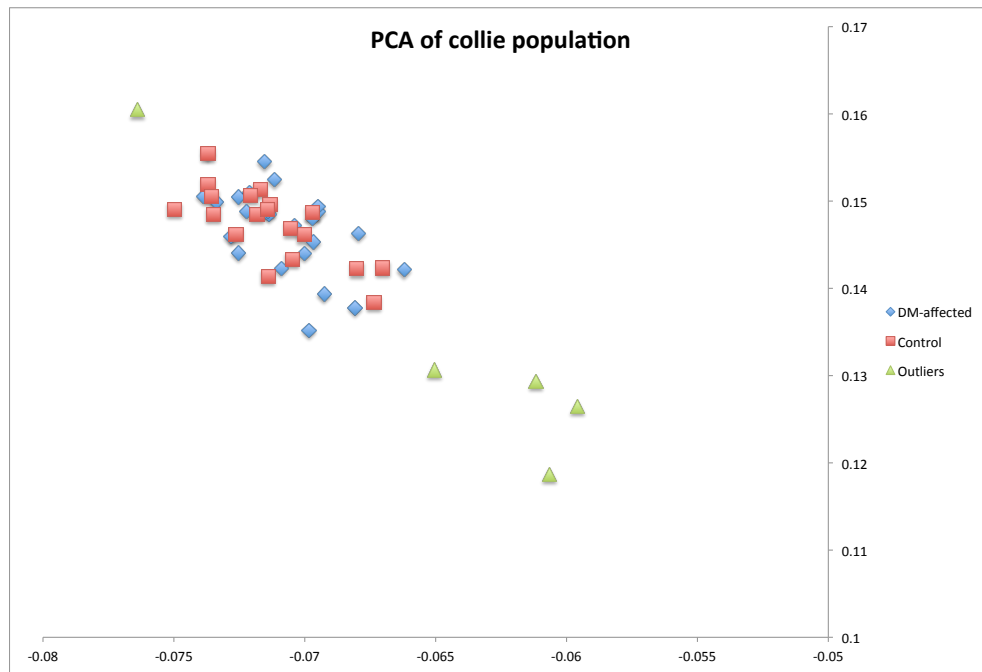
Fig. 1. Principal component analysis of the collie cohort shows 42 dogs clustering together and five outliers. The x-axis is principal component 1 and the y-axis is principal component 2.
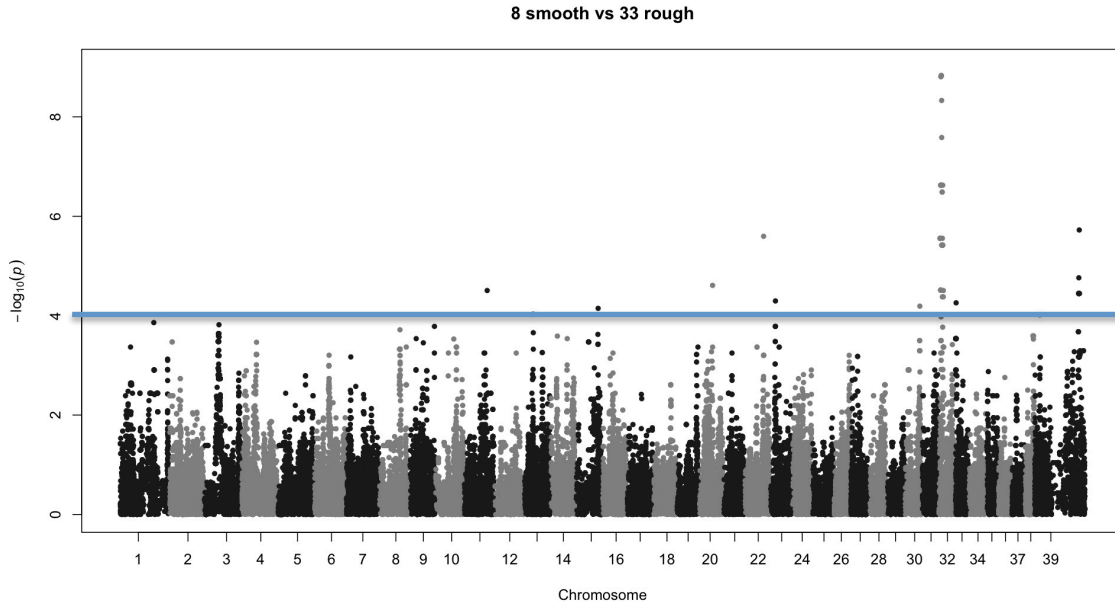
Fig. 2. Manhattan plot showing the results for GWAS for smooth using 173,662 SNPs. The genome-wide *P* values ($-\log_{10} p$) for each SNP are plotted against position on each chromosome. Eight smooth collies versus 33 rough collies show a strong association on chromosome 32. The significance threshold for this plot is $-\log_{10}(p) = 4$.

Two DM-affected collies for which the diagnosis was questionable were excluded from the DM analyses. Case/control analyses were carried out using 21 DM-affected and 19 healthy collies (Fig. 3A). A total of 88 SNPs were significantly associated with DM (*P*raw ≤0.0001), 71 of which were located at the centromeric end of chromosome 10 (Fig. 4). The most significant results obtained were for five SNPs between 2.1 Mb and 2.9 Mb on chromosome 10 (*P*raw = 2.23×10^{-7}). Permutation testing using 100,000 iterations was carried out to reduce any false positive results and the SNPs on chromosome 10 remained most significant (*P*genome = 0.01373). Analysis of genotypes in the chromosome 10 region revealed a 10.5 Mb haplotype for which all DM-affected collies had at least one copy (12 homozygous, nine heterozygous). The haplotype was also present in 74% of

108

healthy collies (two homozygous, 12 heterozygous). There are more than 100 genes in

this 10.5 Mb region.
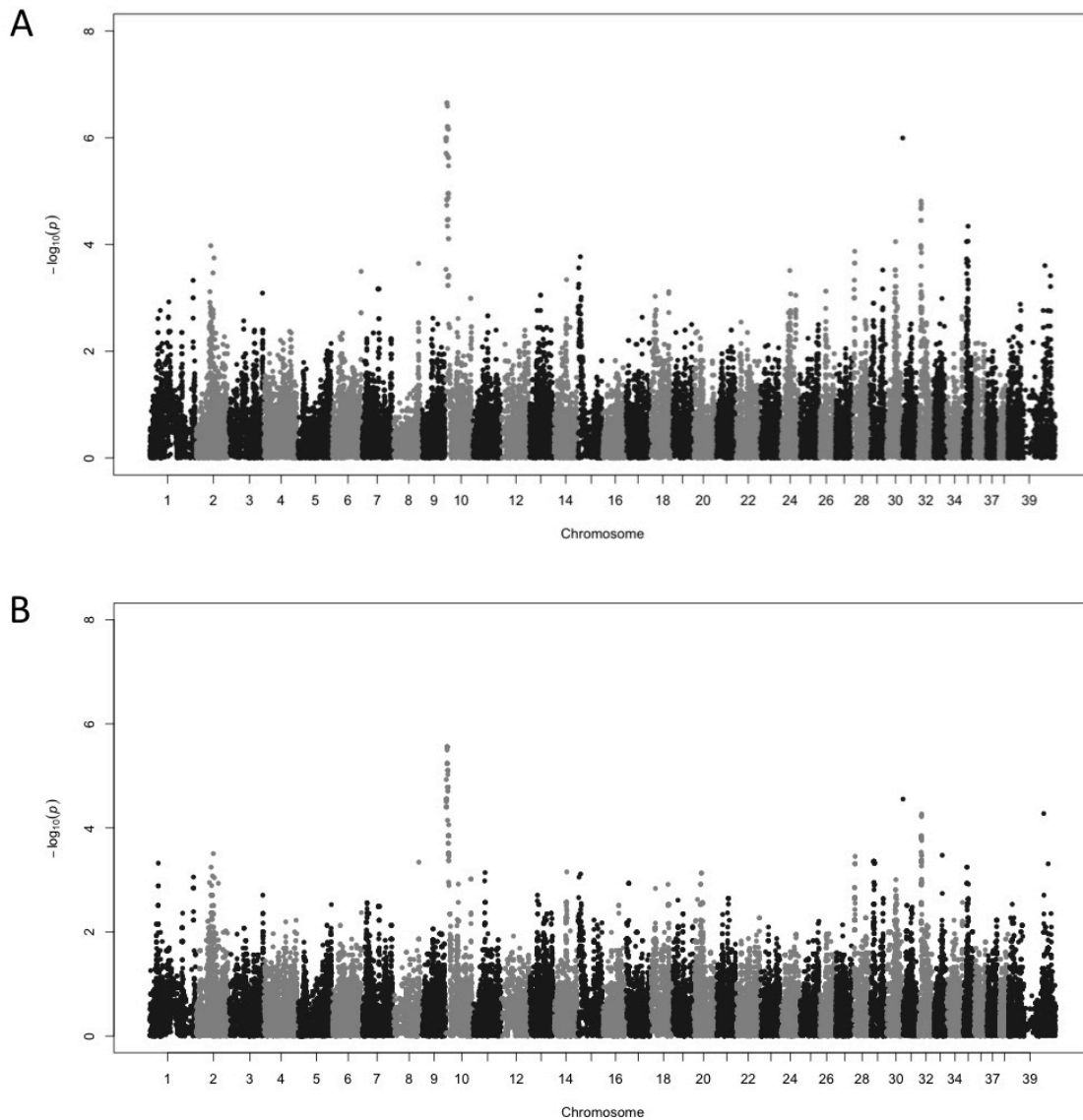
Fig 3. Manhattan plots showing the results for GWAS for DM using 173,662 SNPs. The genome-wide *P* values (-$\log_{10} p$) for each SNP are plotted against position on each chromosome. **A.** Twenty-one DM-affected collies versus 19 healthy collies show a strong association on chromosome 10. **B.** After balancing phenotypes, 17 DM-affected collies versus 14 healthy collies still show a strong association on chromosome 10.
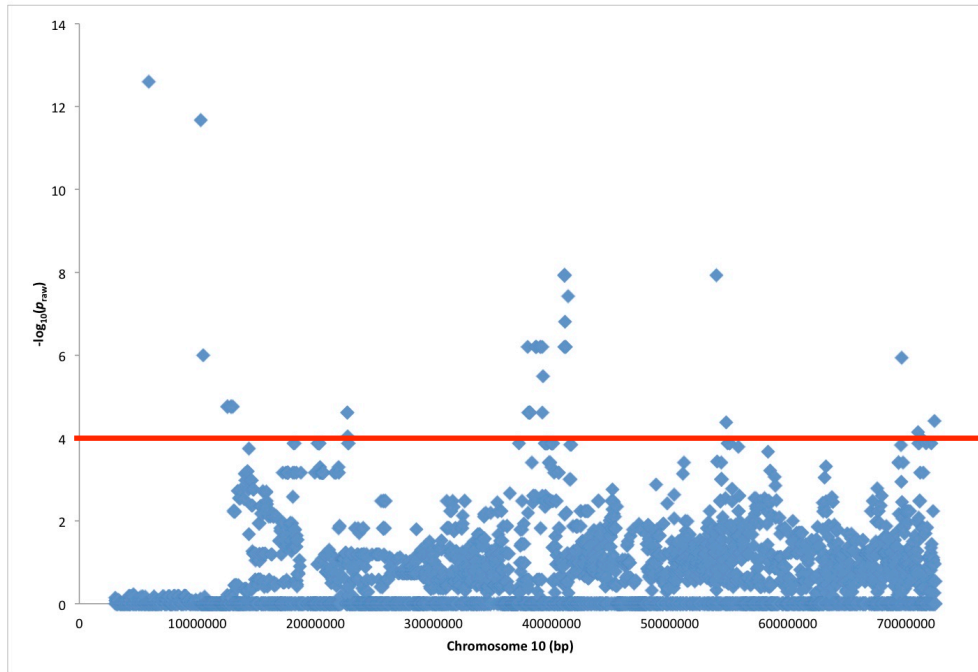
Fig. 4. Scatter plot showing SNP positions along chromosome 10. The raw *P* values for the most strongly associated chromosomes are plotted against chromosomal position for 21 DM-affected collies versus 19 healthy collies. The significance threshold for this plot is $-\log_{10}(p) = 4$.

Of the genes located in this associated DM haplotype, one has a known visible phenotype, *SILV*. This gene is causative for merle coat patterning, a phenotype that segregates in our cohort (30). Merle is autosomal dominant, but dogs having shortened versions of the mutation do not show the merle phenotype (termed cryptic merles). To determine the merle genotype, a direct mutation test was performed for all collies in our study cohort. One DM-affected collie and six healthy controls were determined to be merle, which is a significant difference in distribution (*p*=0.04). One merle healthy control did not exhibit the phenotype. To identify the SNPs associated with merle, a GWAS was performed using seven merle and 35 non-merle collies. A total of 141 SNPs

110

were significantly associated with merle (Praw $\leq$0.0001), 103 of which were located at the centromeric end of chromosome 10. Sixty-five SNPs on chromosome 10 were most significant with a *p*-value of 6.45×10^{-10}. This region is located 2.8 Mb downstream from *SILV*. The five most significant SNPs for DM were not significantly associated with merle (Praw=0.003123).

To confirm that our DM results are not an artifact of association with the non-merle allele in our DM-affected population, the two populations were balanced for merle and a second GWAS was conducted. Five merle collies were removed from healthy control group, leaving one merle dog in each population. Additionally, both populations were balanced for the smooth phenotype. Case/control analyses were rerun for 17 DM-affected (five females/12 males) and 14 controls (eight females/six males) (Fig. 3B). Fifty-six of 68 SNPs significantly associated with DM were on chromosome 10. The eight SNPs most significantly associated with DM were on chromosome 10 between 2.1 Mb and 2.5 Mb (Praw=1.97×10^{-6}). These SNPs were within the haplotype previously identified for DM. There are no mapped genes within this region. Two DM-affected collies were heterozygous for a region defined by the top ten most significant SNPs, while all other DM-affected collies were homozygous. Four SNPs on chromosomes 31, 32, and 39 also reached significance (Praw $\leq$0.0001).

DISCUSSION

In this study, we assembled a cohort of DM-affected and senior, healthy control

111

collies for which we generated whole-genome SNP profiles for the study of DM. With these data, we accurately mapped two coat traits that segregate in our cohort: smooth fur and the merle coat pattern. Analyses to identify genes contributing to DM revealed a primary region of association at the centromeric end of chromosome 10. Associated SNPs span a 10.47 Mb region, and a 10.54 Mb haplotype is common to all affected dogs. The haplotype is larger because it spans from the top of chromosome 10, where there are no SNPs present on the array. The 70 kb region at the top of chromosome 10 can not be excluded because the first SNP on chromosome 10 is within the haplotype for DM and there are no markers upstream to break the haplotype.

Breeding studies have shown DM to be inherited in an autosomal dominant fashion (21). Data generated herein support the hypothesis of an autosomal dominant mode of inheritance because all affected dogs carried either one or two copies of the associated haplotype. There were no correlations between sex or age of onset of DM-affected collies relative to zygosity in the associated region. The presence of the haplotype among healthy controls suggests that the allele may have reduced penetrance. A secondary factor such as an environmental trigger or genetic modifier may be necessary for development of the disease. In human DM, the phenotype results from a combination of environmental and genetic factors (31). Environmental triggers include exposure to toxicodendrons, emotional stress (e.g., the birth of a sibling), and UV radiation (32).

In our study population, we observed an underrepresentation of merles among the

affected collies relative to the controls. The associated region on chromosome 10 encompasses the merle locus, but our data support inheritance of the DM allele on the non-merle chromosome and show that different loci are linked with the two traits.

The number of genes in the DM haplotype region includes five dog RefSeq genes and approximately 130 human RefSeq genes (33,34). The human RefSeq genes are from human chromosome 12 (34). The identified DM haplotype region shows conservation of synteny with human chromosome 12. Several genes have a role in immune function and could be considered candidate genes for DM. Among these are *IL23A*, which has been shown to be upregulated in humans with DM (35), and other cytokines including *IL22* and *IL26*. *IFNG* is also found in this region and has been studied in conjunction with muscle fiber atrophy in DM patients (36). Other genes in this region are involved in the formation of muscle fibers, *MYL6*, and tissue patterning of skin and nerves, *GDF11*.

To narrow the critical interval and reduce the number of candidate genes, future studies will focus on DM in other breeds, primarily Shetland sheepdogs. Collies and Shetland sheepdogs share a common ancestor in the early herding dog of Scotland and have been interbred as recently as the 1950s. It is likely that they share the genetic basis for DM. An across-breed approach will allow us to take advantage of the differences in homogeneity that exists between the breeds (37). Specifically, by comparing the DM haplotype region identified in affected collies to the same region in affected Shetland sheepdogs will elucidate differences between the two breeds and refine the critical

113

interval. Upon refinement of the critical interval, positional candidate genes will be identified and investigated for candidate causal mutations.

In conclusion, we used genome-wide association and haplotype analysis to localize DM in the collie to canine chromosome 10. There are 130 genes in this region and future work will focus on an across-breed approach to narrow down the critical interval.

REFERENCES

1. Ramanan AV, Feldman BM (2002) Clinical features and outcomes of juvenile dermatomyositis and other childhood onset myositis syndromes. Rheum Dis Clin North Am 28:833-857.

2. McCann LJ, Juggins AD, Maillard SM, Wedderburn LR, Davidson JE et al. (2006) The Juvenile Dermatomyositis National Registry and Repository (UK and Ireland)--clinical characteristics of children recruited within the first 5 yr. Rheumatology (Oxford) 45:1255-1260.

3. Symmons DP, Sills JA, Davis SM (1995) The incidence of juvenile dermatomyositis: results from a nation-wide study. Br J Rheumatol 34:732-736.

4. Mendez EP, Lipton R, Ramsey-Goldman R, Roettcher P, Bowyer S et al. (2003) US incidence of juvenile dermatomyositis, 1995-1998: results from the National Institute of Arthritis and Musculoskeletal and Skin Diseases Registry. Arthritis Rheum 49:300-305.

5. Dimachkie MM (2011) Idiopathic inflammatory myopathies. J Neuroimmunol 231:32-42.

6. Bohan A, Peter JB (1975) Polymyositis and dermatomyositis (first of two parts). N Engl J Med 292:344-347.

7. Bohan A, Peter JB (1975) Polymyositis and dermatomyositis (second of two parts). N Engl J Med 292:403-407.

8. Feldman BM, Rider LG, Reed AM, Pachman LM (2008) Juvenile dermatomyositis and other idiopathic inflammatory myopathies of childhood. Lancet 371:2201-2212.

9. Bowyer SL, Blane CE, Sullivan DB, Cassidy JT (1983) Childhood dermatomyositis: factors predicting functional outcome and development of dystrophic calcification. J Pediatr 103:882-888.

10. Pachman LM, Hayford JR, Chung A, Daugherty CA, Pallansch MA et al. (1998) Juvenile dermatomyositis at diagnosis: clinical characteristics of 79 children. J Rheumatol 25:1198-1204.

11. Spencer CH, Hanson V, Singsen BH, Bernstein BH, Kornreich HK et al. (1984) Course of treated juvenile dermatomyositis. J Pediatr 105:399-408.

12. Huber AM, Lang B, LeBlanc CM, Birdi N, Bolaria RK et al. (2000) Medium- and long-term functional outcomes in a multicenter cohort of children with juvenile dermatomyositis. Arthritis Rheum 43:541-549.

13 .Constantin T, Ponyi A, Orbán I, Molnár K, Dérfalvi B et al. (2006). National registry of patients with juvenile idiopathic inflammatory myopathies in Hungary-clinical characteristics and disease course of 44 patients with juvenile dermatomyositis. Autoimmunity 39:223-232.

14. Ben-Basser M, Machtev I (1972) Picornavirus like structures in acute dermatomyositis. Am J Clin Pathol 58:245.

15. Travers RL, Hughes GRV, Cambridge G, Sewell JR (1977) Coxsackie B neutralisation titres in polymyositis/dermatomyositis. Lancet 1:1268.

16. Pittsley RA, Shearn MA, Kaufmann MD (1978) Acute hepatitis B simulating dermatomyositis. J Amer Med Assoc 239:959.

17. Christenson ML, Pachman LM, Schneiderman R, Patel DC, Friedman JM (1986)

116

Prevalence of coxsackie B virus antibodies in patients with juvenile dermatomyositis. Arthritis Rheum 29:1365.

18. Manlhiot C, Liang L, Tran D, Bitnun A, Tyrell PN et al. (2008) Assessment of an infectious disease history preceding juvenile dermatomyositis symptom onset. Rheumatology 47:526-529.

19. Cooper GS, Miller FW, Pandey JP (1999) The role of genetic factors in autoimmune disease: Implications for environmental research. Environ Health Perspect 107:693-700.

20. Rees CA, Boothe DM, Wilkie S (2002) Therapeutic response to pentoxifylline and its active metabolites in dogs with dermatomyositis. Veterinary Dermatology 13: 211-229.

21. Hargis AM, Haupt KH, Hegreberg GA, Prieur DJ, Moore MP (1984) Familial canine dermatomyositis: Initial characterization of cutaneous and muscular lesions. The American Journal of Pathology 116: 234-244.

22. Hargis AM, Haupt KH, Prieur DJ, Moore MP (1985) A skin disorder in three Shetland sheepdogs: a comparison with familial canine dermatomyositis in collies. The Compendium on Continuing Education for the Practicing Veterinarian 4: 306-315.

23. Hargis AM, Prieur DJ, Haupt KH, Collier LL (1986) Postmortem findings in a Shetland sheepdog with dermatomyositis. Veterinarian Pathology 23: 509-511.

24. Collie and Shetland sheepdog (2011) Breed description from the AKC website. Retrieved from http://www.akc.org/breeds/ on 3 April 2011.

25. Awano T, Johnson GS, Wade CM, Katz ML, Johnson GC et al. (2009) Genome-wide association analysis reveals a SOD1 mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis. Proc Natl Acad Sci U S A 106:2794-2799.

26. Gill JL, Tsai KL, Krey C, Noorai RE, Vanbellinghen J et al. (2012) A canine BCAN microdeletion associated with Episodic Falling Syndrome. Neurobiology of Disease 45:130-136.

27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575.

28. R Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

29. Housley DJ, Venta PJ (2006) The long and the short of it: evidence that FGF5 is a major determinant of canine 'hair'-itability. Anim Genet 37:309-315.

30. Clark LA, Wahl JM, Rees CA, Murphy KE (2006) Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. Proc Natl Acad Sci U S A. 103:1376-1381.

31. Shah M, Mamyrova G, Targoff IN, Huber AM, Malley JD et al. (2013). The clinical phenotypes of the juvenile idiopathic inflammatory myopathies. Medicine 92:25-41.

32. Rider L, Miller FW (2011) Deciphering the clinical presentations, pathogenesis, and treatment of the idiopathic inflammatory myopathies. JAMA 305:183-190.

33. Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res Apr

12(4):656-664.

34. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH et al. (2002) The human genome browser at UCSC. Genome Res Jun 12(6):996-1006.

35. Shen H, Xia L, Lu J, Xiao W (2011) Interleukin-17 and interleukin-23 in patients with polymyositis and dermatomyositis. Scand J Rheumatol May;40(3):217-20.

36. Gallardo, E, de Andres, I, Illa, I (2001) Cathepsins are upregulated by IFN-gamma/STAT1 in human muscle culture: a possible active factor in dermatomyositis. Journal of Neuropathol Exp Neurol. 60(9):847-855.

37. Sutter NB, Ostrander EA (2004) Dog star rising: the canine genetic system. Nat Rev Genet. 5:900-910.

119

CHAPTER VI

WHOLE-GENOME RESEQUENCING OF THE COLLIE FOR DISCOVERY OF

GENOMIC VARIATIONS

Rooksana E. Noorai[a], Alison Starr-Moss[a], Margaret E. Staton[b], Leigh Anne Clark[a]

[a] Department of Genetics and Biochemistry, College of Agriculture, Forestry and Life

Sciences, Clemson University, Clemson, SC 29634, USA

[b] Clemson University Genomics Institute, Clemson University, Clemson, SC 29634, USA

Unpublished

ABSTRACT

Domestic dogs have been under artificial selection by humans for hundreds of years. *Canis lupus familiaris* is unmatched in phenotypic diversity, compared to other mammals, and is an excellent model for investigating the genetic basis of phenotypic diversity, due to their various physiology, morphology, and behavior traits. We resequenced the genome of five American collies to 113-fold coverage using Illumina HiSeq paired-end sequencing. Trimmed and concordant reads were mapped to the canine reference genome (Broad CanFam3.1), generated from a boxer. We identified more than 9.7 million single nucleotide polymorphisms (SNPs), which showed concordance rates >98% with Illumina BeadChip SNP data. The predicted consequence of each SNP was determined based on its position relative to mapped genes. We also identified 671,197 structural variants including deletions, insertions, inversions, tandem duplications, and break points shared by all 5 collies, relative to the canine reference genome. Genetic regions, with significantly reduced heterogeneity surrounding a gene(s) under strong positive selection, were detected on 9 different chromosomes. Candidate genes for two phenotypes (Irish spotting and dolichocephaly) fixed in the collie breed were identified. Sequence data generated herein provide a resource for the identification of variants responsible for morphological traits and heritable diseases of the collie.

INTRODUCTION

The dog, *Canis lupus familiaris,* has been a best friend to mankind since their domestication more than 14,000 years ago (1). Through artificial selection, man has created more than 400 different pure breeding lines, each arising from a need for certain attributes (2). Selected traits include behaviors like herding, guarding, hunting abilities, and morphological features like body size, coat texture, and ear carriage. As a result, dogs are the most phenotypically diverse species among mammals. Purebred dogs are maintained through closed breeding populations. Levels of homogeneity within breeds are very high, while levels between breeds is considerably lower (3). Human populations have high rates of heterogeneity and thus require large numbers of individuals to map hereditary disorders (4,5). For this reason, purebred dog populations offer great models for the study of hereditary diseases in humans.

Because dog breeds exhibit as many as 340 naturally occurring analogous human diseases (6), the NIH prioritized the dog genome for whole-genome sequencing. In 2005, the first draft of the canine reference sequence was released: a 7.6 X coverage, high-quality assembly generated for a female boxer (2*)*. The estimated sequence length, including gaps, is 2.5 Gb. Roughly 3% of the sequence assembly was highly repetitive and could not be mapped (2). The reference genome covers more than 98% of the euchromatic regions (2*)*. This consortium also generated a dense SNP map, using 3 different cohorts, that consists of more than 2.5 million single nucleotide polymorphisms (SNPs) (2). First, the consortium identified within the boxer sequence SNPs having a

122

different allele on two or more unique reads (~770,000 SNPs). Second, they compared the boxer draft sequence to the previously published 1.5 X standard poodle sequence (~1,460,000 SNPs) (7). Third, shotgun sequence data for nine various dog breeds (~100,000 reads each, 0.02-fold coverage), four grey wolves and one coyote (~22,000 reads each, 0.004-fold coverage) was aligned to the boxer (~440,000 SNPs). There have been a total of three different versions (builds) of the canine reference genome. The first build was assembled using a version of the ARACHNE program (8) and is referred to as CanFam1.0; most of the initial quality analyses of the genome were based on this version (2). A second build (CanFam2.0) is an updated assembly that was released in 2005 with minimal improvements over CanFam1.0 (2). Currently, canine SNP arrays (Affymetrix and Illumina brand) are based on the CanFam2.0 build of the reference genome. In 2012, a third build was released (Broad CanFam3.1) with improvements including: an N50 (the measure of a majority of the contig lengths in an assembly, 9) ~1.5 times longer, ~8,000 fewer contigs, and an 8-fold reduction of assembled sequence in gaps (10). The size of the canine reference genome in Broad CanFam3.1 is 2.4 Gb (2).

New methods of whole-genome sequencing technologies were introduced in the beginning of the 21st century that allowed for massive parallelization. This high-throughput method, termed next-generation sequencing (NGS), greatly reduced the cost and time for generating data. Illumina's HiSeq technology simultaneously sequences millions of shorter read fragments (11). These short reads can be mapped back to a reference genome, but are not ideal for generating a *de novo* assembly sequence because

123

of the high incidence of misassembling regions containing repeats and the high redundancy of the short reads (12).

The whole-genome sequencing projects of three dog breeds have been published, a 1.5X coverage of the poodle, a 7.6X coverage of the Boxer, and a 45X coverage of the Korean Jindo (poodle, boxer, Korean Jindo references). Herein, we resequenced the complete genomes of 5 collies and mapped the data to the CanFam3.1 build of the canine reference genome. We identified SNPs and structural variants (insertions, deletions, tandem duplications, inversions, and breakpoints) present in the collie and not the boxer. We also verified concordance of the SNPs identified in the whole-genome resequencing by running Illumina Canine HD BeadChip arrays on the same collies. We then classified the effect of the genetic variants we identified (13). Furthermore, we identified 9 chromosomes with regions of reduced heterogeneity and propose two candidate genes for selection in the collie.

MATERIALS AND METHODS

Sample Collection

Collie owners were contacted via electronic methods to solicit research participants. Pedigrees, information regarding age, sex, coat color, and hair length, and medical histories were collected from purebred American collies. Whole blood for DNA was collected by the primary care veterinarian of each dog. In order to maximize data gleaned from the genome sequences, pedigrees were used to select the five most

124

genetically diverse American collies. Two dogs shared a common ancestor in the fourth generation while no common ancestors were present in the first six generations of the others. Among the dogs selected for sequencing were males and females, roughs and smooths, and solids and blue merles (Table 1).

Whole-genome sequencing

Genomic DNA was isolated using the DNeasy blood and tissue kit (QIAGEN, Valencia, USA) and adjusted to a concentration of 50 ng/µL in 10mM Tris-CL buffer pH 8.5, measured by optical density using a NanoDrop 1000 Spectrophotometer (Thermo Scientific, Wilmington, USA). One hundred fifty ng of DNA was used for 2x100 bp paired-end sequencing, on an Illumina HiSeq 2000 at the University of Missouri DNA Core Facility (Columbia, Missouri, USA). Two compressed fastq files were generated per lane containing each end of the pair.

Data preparation

The quality of each set of reads was examined using FastQC v0.10.1 (14). Trimming and sorting of all reads was conducted using Trimmomatic v0.30 (15). The output of forward and reverse, paired and unpaired reads were re-examined to verify an improvement of quality. Some of the metrics used to assess quality include: improvement in the per base sequence quality, the per sequence quality scores, the absence of overrepresented sequences, and low kmer content (15).

125

The Broad CanFam3.1 "soft-masked" assembly sequence (2) was downloaded from the UCSC Genome Bioinformatics site (13) and formatted for indexing in Bowtie2 v2.1.0 (16). "Soft-masked" refers to the representation of repeat sequences as capital letters and all other sequences as lower case letters. Bowtie2 was used to align all paired forward and reverse reads per lane to the Broad CanFam3.1 genome. SAMtools v0.1.18 (17) was used to convert the Bowtie2 output alignments for sorting and indexing. Interactive Genomics Viewer v2.3 was used to visualize the results (18,19).

DATA ANALYSIS

Coverage, single nucleotide polymorphism and structural variant discovery

GenomeCoverageBed (2009), a part of the bedTools suite v 2.16.2 (20), was used to generate data in order to calculate the average sequence coverage for each collie. A variant call file (VCF) v4.1 combining all five collies was produced using SAMtools (17) and BCFtools (17).

The sorted and indexed result files for all five collies were used as input to Pindel v0.2.4t (21) and structural variants (deletions, inversions, tandem duplications, short insertions, long insertions, and unassigned breakpoints) relative to the reference genome were detected. Structural variants present in all five collies were identified.

The Variant Effect Predictor script release 73 by Ensembl (22) was run on our complete VCF containing data from all five collies to categorize the location and/or type of amino acid or splice site change predicted by each SNP.

Selective sweeps

To identify selective sweeps, we implemented the creeping window method described by Qanbari *et al.* (23). We identified all SNPs in our whole-genome sequence, outside of "soft-masked" regions, and used a window size no larger than 1 Mb. We ignored gaps of greater than 10 kb between SNPs. If a creeping window had 50 or fewer SNPs, then it was not considered in our data set. The heterozygosity (Hp) statistic was calculated for all creeping windows. All Hp values were averaged and the standard deviation was calculated. Each Hp value was Z-transformed, making the average Hp value 0 and the standard deviation equal to 1.

Concordance rate between whole-genome sequence and Illumina arrays

Genomic DNA from each collie was adjusted to a concentration of 100 ng/μL. Illumina CanineHD Infinium BeadChips were used to profile 173,662 SNP genotypes per collie, across autosomal chromosomes 1 through 38 and sex chromosomes 39/X and 40/Y. The BeadChips were processed at the Cornell University Life Sciences Core Laboratories Center (Ithaca, New York, USA) according to manufacturer's protocols. Raw data files were analyzed using GenomeStudio's (2011.1) Genotyping Module v1.9 to generate SNP calls. The PLINK Input Report Plug-in v2.1.1 was used to format the

127

data. BeadChip SNP positions were remapped to the Broad CanFam3.1 using the UCSC Genome Browser's LiftOver script (24) to determine the concordance rate with the whole-genome resequence data. Array SNPs with missing data were dropped from the comparison.

RESULTS

Study Cohort

In order to identify the most genetically diverse collies for whole-genome sequencing, pedigrees were analyzed to determine relationships. Five genetically diverse American collies were selected. Two dogs shared a common ancestor in the fourth generation with no common ancestors present in the first six generations amongst the others. Among the dogs selected for sequencing were males and females, roughs and smooths, and solids and blue merles (Table 1).

Collie	Sex	Coat Type	Coat Pattern	Age at Collection (years)	Inherited Autoimmune Disease
1	M	Smooth	Solid	1	Yes
2	F	Rough	Merle	10	No
3	M	Rough	Merle	9	No
4	M	Rough	Solid	< 1	Yes
5	F	Rough	Solid	7	Yes

Table 1. Collies selected for whole-genome sequencing.

128

Whole-genome resequencing

Genomic DNA from five collies was sequenced using the Illumina HiSeq 2000 with individual paired-end libraries. The following fragments were gel size selected: Collie 1 – 478 bp, Collie 2 – 530 bp, Collie 3 – 467 bp, Collie 4 – 526 bp and Collie 5 – 513 bp. The range of the pass filtering rate was from 75% - 93%. The number of reads produced ranged from 531 to 997 million, providing roughly 48 to 75 Gb of sequence (see Supplementary Table 1).

Sequence Analysis

Paired-end reads were trimmed for low quality scores in the leading 3 bp, trailing 6 bp, and within a sliding window of size 4 bp. Reads were eliminated if their minimum length fell below 36 bp. Reads were mapped to the canine reference genome that was "soft-masked" for repeats (Broad CanFam3.1/canFam3) using Bowtie2 with sensitive parameters. Genome coverage ranged from 16.8 X – 24.9 X, resulting in a combined 113-fold coverage.

SNP Detection

Putative SNPs were determined by comparing the reference sequence to the aligned reads of all five collies. More than 9.7 million unique SNPs were identified across the genome, including unmapped linkage groups. Genotype data from Illumina Canine HD beadchips were used to validate a subset of the identified SNPs. A total of 167,692 SNPs were placed on the reference genome (11,740 SNPs from the SNP array

could not be remapped to Broad CanFam3.1 and were discarded from further analysis).

Concordance rates ranged from 98% to over 99% between the BeadChip genotypes and

whole-genome resequencing genotypes (Table 2). While this method does not validate all

identified SNPs, the SNPs it does validate were identified as highly informative (25).

	SNPs total	SNPs same	SNPs different	SNPs NA	Concordance
Collie 1	29,755	29,656	99	525	99.67%
Collie 2	29,757	29,647	110	523	99.63%
Collie 3	29,688	29,201	487	592	98.36%
Collie 4	29,762	29,675	87	518	99.71%
Collie 5	29,746	29,642	104	534	99.65%

Table 2. Comparison of Illumina CanineHD BeadChip genotypes to sequencing SNPs. These results indicate a low false positive rate of SNP discovery. NA genotypes result from missing results from either method for a given SNP.

The Variant Effect Predictor, available through Ensembl, was used to assess and

categorize the location of all SNPs and small variations, compiled in the VCF file,

relative to mapped genes present on chromosomes 1 – 38 and X. SNPs were

characterized based on location (intergenic, intronic, upstream, downstream, splice site,

5' UTR, or 3' UTR) or type of change (frameshift, missense, synonymous, stop gained,

stop lost, stop retained, or within mature miRNA). Detailed results can be seen in Table

3.

Functional class	collies	% of total
Intergenic	7,623,282	61.1
Intronic	3,607,150	28.9
Upstream	604,082	4.8
Downstream	512,454	4.1
3' UTR	47,607	0.4
Missense	23,258	0.2
Synonymous	22,571	0.2
5' UTR	16,312	0.1
Splice site	9,688	0.1
Within non coding gene	4,220	0.0
Frameshift	2,798	0.0
Stop gained	602	0.0
Within mature miRNA	189	0.0
Stop lost	58	0.0
Total	12,474,271	100.0

Table 3. SNP Functional class membership for collies. SNPs were identified by sequencing of five collie genomes and grouped by predicted functional consequences. The percentage of SNPs in each class are in the last column.

Structural Variants Detection

A total of 4,298,455 structural variants were detected using Pindel (21) including: deletions, inversions, short insertions, tandem duplications, long insertions, and other unassigned break points (Table 4). A total of 675,341 variants were found to exist in all five collies across the genome, including unmapped linkage groups. The remaining 3.6 million structural variants were not common to all five collies.

	Deletions	Inversions	Short Insertions	Tandem Duplication	Long Insertions	Other Break Points	Total SV
Totals	1383310	46174	952650	11021	952650	952650	**4298455**
5 collies - whole genome w/ unmapped linkage groups	357642	1771	260979	1554	3533	49862	**675341**
chr1	19037	105	13708	91	200	3071	**36212**
chr2	12513	71	8630	63	150	1874	**23301**
chr3	14553	60	10691	53	122	1849	**27328**
chr4	13917	65	9969	59	127	1644	**25781**
chr5	13790	58	9731	56	123	2184	**25942**
chr6	10022	53	7182	46	106	1701	**19110**
chr7	9954	44	7381	41	99	1377	**18896**
chr8	13128	62	9507	55	138	1803	**24693**
chr9	8680	69	6085	34	78	1687	**16633**
chr10	11426	45	8156	56	110	1834	**21627**
chr11	10372	43	7536	48	88	1383	**19470**
chr12	14686	75	10782	68	133	1710	**27454**
chr13	8280	42	6141	30	80	1069	**15642**
chr14	8591	31	6618	36	73	951	**16300**
chr15	11557	54	8588	41	112	1573	**21925**
chr16	8692	33	6432	39	87	1134	**16417**
chr17	9359	48	6917	49	89	1302	**17764**
chr18	8734	48	6149	56	89	1312	**16388**
chr19	9647	41	7269	34	89	1073	**18153**
chr20	6907	36	4797	26	71	1310	**13147**
chr21	7470	35	5340	29	61	1045	**13980**
chr22	8680	31	6468	27	80	1039	**16325**
chr23	8865	51	6873	30	89	1061	**16969**
chr24	7457	35	5306	38	83	1118	**14037**
chr25	9511	37	7138	30	99	1175	**17990**
chr26	6198	57	4366	26	72	999	**11718**
chr27	6641	32	4691	29	59	907	**12359**
chr28	6130	32	4323	23	59	934	**11501**
chr29	8318	37	6370	29	75	900	**15729**
chr30	6637	35	4929	24	58	965	**12648**
chr31	6896	37	5209	20	60	776	**12998**
chr32	7973	35	5953	34	61	708	**14764**
chr33	5142	24	3807	24	47	676	**9720**
chr34	8949	47	6670	33	83	1050	**16832**
chr35	5189	25	3927	20	61	635	**9857**
chr36	5602	32	4243	20	53	633	**10583**
chr37	5526	28	4232	28	39	818	**10671**
chr38	3834	26	2810	22	38	591	**7321**
chrX	6899	37	4783	41	97	1155	**13012**
Total	**355762**	**1756**	**259707**	**1508**	**3438**	**49026**	**671197**

Table 4. Results from Pindel analyses. All deletions, inversions, short insertions, tandem duplications, long insertions, and other break points were identified across the reference genome and existing in all five sequenced collies. Totals include all mapped chromosomes and linkage groups, as well as all chromosomes individually.

We identified 671,197 structural variants on chromosomes 1 – 38 and X that were present in all five collies. The remaining 4,144 variants were found in the unmapped linkage groups.

There were 355,762 deletions shared by all five collies. Deletions varied in length from one to 32,787 bp, with the average being 40 bp and the median deletion being two bp. There were 1,756 inversions between one and 32,887 bp long; the average and median inversion of 16,897 bp. Tandem duplications varied from 29 to 32,776 bp, with an average and median of 22,160 bp. A total of 259,707 short insertions varied from one to 77 bp in length and had an average and median of one bp.

Selective Sweep Discovery

We identified regions of pooled heterozygosity (H_p), in order to detect signals of recent selection. Z-transformation of the H_p values resulted in a mean value of 0 and a standard deviation of 1. The negative ZH_p distribution was plotted to show putatively selected regions greater than 3.4 standard deviations from the mean (Fig. 1). A region with significantly reduced heterogeneity, surrounding a gene(s) being selected for, is called a selective sweep. Based on our data, we hypothesize the existence of selective sweeps on chromosomes 1, 5, 8, 10, 12, 15, 24, 29, and 34. The SNPs used to demarcate the regions were not part of repeat regions.
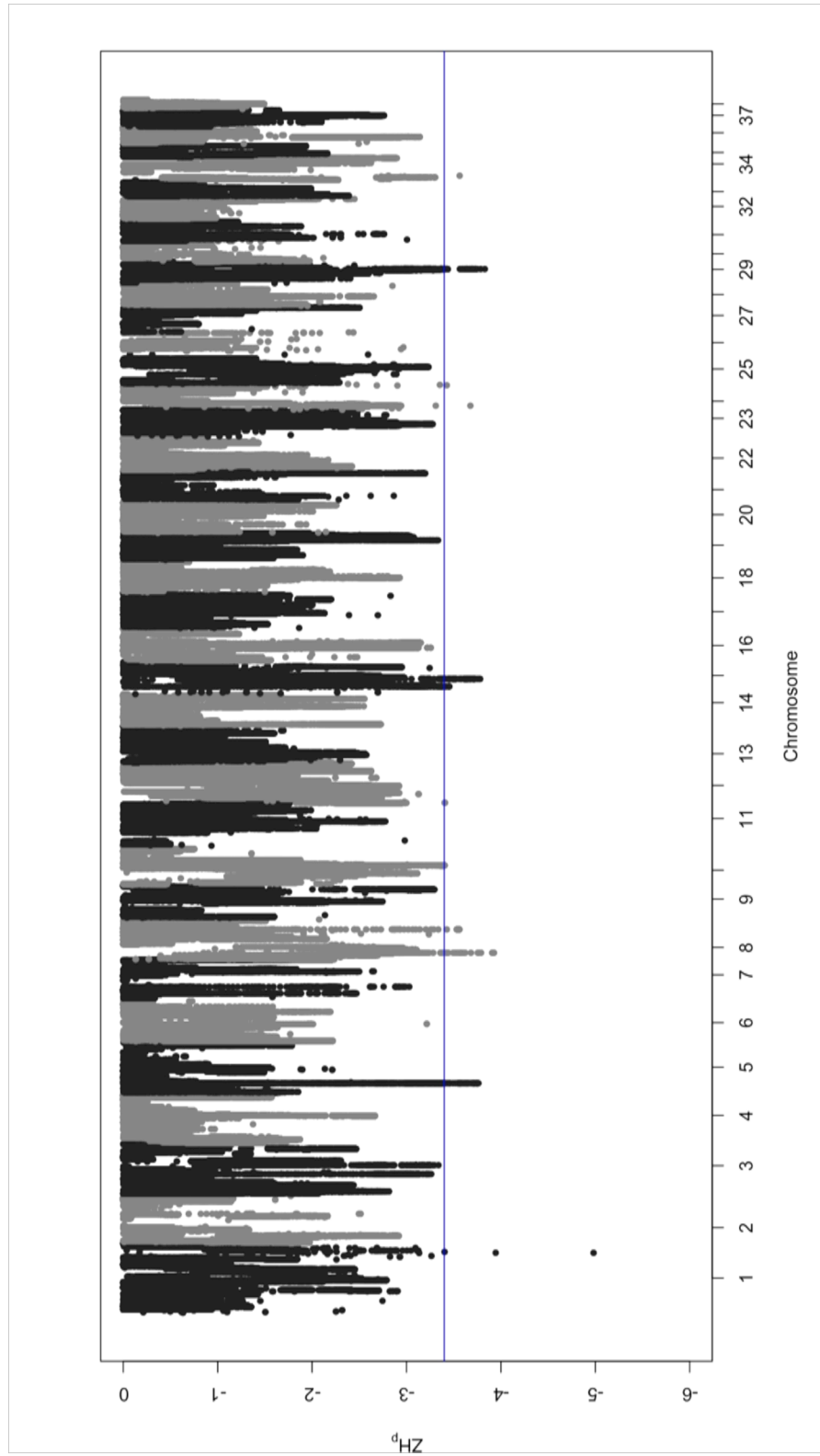
Fig. 1. The negative tail of the ZH_p distribution presented along autosomes 1-38. The X chromosome was not included because the sex of the five collies was not considered. Each dot represents a creeping window of up to 1 Mb. The horizontal blue line stands for the significance level at $ZH_p = -3.4$.

DISCUSSION

In this study, we used an Illumina HiSeq 2000 sequencer and paired-end reads to obtain mapped genome sequence of the collie. The high coverage sequence data generated herein provides a novel resource for future studies in the breed. Previous work from our laboratory indicated that SNPs identified from the canine reference sequence project were largely uninformative in collies, presenting a challenge for the mapping of phenotypes. In addition to SNPs, our data revealed 671,197 structural variants (deletions, insertions, inversions, tandem duplications, and break points) present in all 5 collies.

We identified over 9.7 million SNPs relative to the canine reference genome. This number is comparable to that reported by a recent dog resequencing project that reported more than 4.6 million SNPs detected from a single male dog (28). The high number of SNPs observed in the single Korean Jindo dog, compared this study of five collie dogs, may be explained by the Korean Jindo breed being more divergent from the boxer or by mapping error in the Korean Jindo's aligned sequence (28). Further indication that the collie exhibits increased homozygosity is provided by a large number of chromosomes having possible selective sweeps. A total of 9 chromosomes whose loss of heterozygosity would indicate fixation of long stretches of DNA were identified.

Two chromosomes, CFA12 and 15, were of specific interest because they harbor dog leukocyte antigen (DLA) genes and *KITLG*, respectively. The significant reduction of heterogeneity surrounding these genes suggests the presence of selective sweeps. DLA

135

is important in immune recognition and regulation (29,30). *KITLG* encodes the ligand of the tyrosine-kinase receptor and harbors mutations in other species that affect melanogenesis (31). All collies are fixed for a coat pattern termed Irish spotting, which is characterized by a white collar around the neck, white feet, and a white tail tip. Although no coding mutations were detected in *KITLG*, two 200-300 bp deletions in the 5' region are homozygous in all five collies and will be further investigated as candidate mutations for Irish spotting.

A selective sweep on CFA 24 encompasses *BMP7*, a gene important in skeletal patterning (32). Bone and cartilage result from the transformation of mesenchymal cells with a protein encoded by *BMP7*. This gene may be involved in the head morphology of collies, characterized by distinct dolichocephaly – extremely long muzzles relative to the width of their heads. Other chromosomes displayed comparable levels of sweeps and analysis of these regions is ongoing.

A large number of structural variants (deletions, insertions, tandem repeats, inversions, and unassigned breakpoints) were observed among the collie genomes. The length of some variants was extremely long, and further investigation of individual variants revealed the presence of false positives. Because the software, Pindel (21), analyzing the data had no threshold set regarding a minimum number of reads, if only one read in each of the collies matched the variant, then it was called. This shortcoming was most notable among the deletion class of variants. Several extremely large deletions

were present in all five dogs, but were not supported by a majority of the reads in the regions. Without the support of approximately 50% of the total reads per dog, these deletions are erroneous. Future work should include additional analyses to increase the stringency of read thresholds.

The very high concordance rates between the whole-genome resequencing data and the BeadChips confirm that both data sets are high-quality. By extension, the unvalidated SNPs are considered to be of the same quality. Of the whole-genome resequencing SNPs that were discarded, 13 were from the Y chromosome, which is not covered by the canine reference sequence. The variant predictor software by Ensembl identified any functional effect changes resulting from SNPs identified in the collie genome. While many SNPs were predicted to result in missense, nonsense, and frameshift mutations, these data are based on predicted coding regions and do not account for unmapped genes or errors in the annotation. The investigation and validation of these variants is beyond the scope of this study.

In conclusion, we generated 113X coverage resequencing data of the collie genome. Additional data are necessary to determine the significance of the structural variants and the selective sweeps identified herein.

REFERENCES

1. Moody JA, Clark LA, Murphy KE (2006) Breed Clubs and Canine History. In The Dog and its Genome. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press pp1-18.

2. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog Nature Dec 8 438(7069):803-19.

3. Parker HG, Sutter NB, Ostrander EA (2006) Understanding Genetic Relationships among Purebred Dogs: The PhyDo Project. In The Dog and its Genome. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press pp.141-158.

4. Kidambi S, Ghosh S, Kotchen JM, Gim CE, Krishnaswami S et al. (2012). Non-replication study of a genome-wide association study for hypertension and blood pressure in African Americans. BMC Medical Genetics. 13:27.

5. Tanikawa C, Urabe Y, Matsuo K, Kubo M, Takahashi A et al. (2012). Nature Genetics. 44(4): 430-4.

6. The University of Sydney (2013, Nov 15) OMIA – Online Mendelian Inheritance in Animals Retrieved November 15, 2013 from http://omia.angis.org.au/home/.

7. Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K et al. (2003) The dog genome: survey sequencing and comparative analysis. Science 301:1898-1903.

8. Jaffe DB, Butler J, Gnerre S, Mauceli E, Lindblad-Toh K et al. (2003) Whole-genome sequence assembly for mammalian genomes: Arachne 2. Genome Research 13**:** 91–96.

9. National Center for Biotechnology (2008, Jan. 6) Commonly Used Genome Terms Retrieved November 15, 2013 from http://www.ncbi.nlm.nih.gov/projects/genome/glossary.shtml/.

10. NCBI (2011, Nov. 2) CanFam3.1 – Assembly. *NCBI* Retreieved September 4, 2013, from http://www.ncbi.nlm.nih.gov/assembly/GCF_000002285.3/.

11. Illumina (2013) A sequencer for every need. Every Budget. Every lab Retrieved on November 15, 2013 from http://www.illumina.com/systems/sequencing.ilmn/.

12. Illumina (2010) Technical Note: Sequencing. De Novo Assembly Using Illumina Reads Retrieved November 15, 2013 from http://res.illumina.com/documents/products/technotes/technote_denovo_assembly_ecoli.pdf/.

13. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH et al. (2002) The human genome browser at UCSC. Genome Res Jun 12 (6):996-1006.

14. Babragan Bioinformatics (2012, Mar. 5) FastQC A Quality Control Tool for High Throughput Sequence Data. *FastQC* Retrieved September 4, 2013, from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

15. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE et al. (2012) A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res Jul 40(Web Server issue):W622-7.

16. Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods. 9:357-359.

17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al. (2009). The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25:2078-9.

18. Thorvaldsódttir H, Robinson JT, Mesirov JP (2012) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 2012.

19. Robinson JT, Thirvaldsdóttir H, Winckler W, Guttman M, Lander ES et al. (2011) Integrative Genomics Viewer. Nature Biotechnology 29:24-26.

20. Quinlan AR, Hall IM, (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6):841–842.

21. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 25(21): 2865-2871.

22. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P et al. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. BMC Bioinformatics 26(16):2069-70.

23. Qanbari S, Strom TM, Haberer G, Weigend S, Gheyas AA et al. (2012) A High Resolution Genome-Wide Scan for Significant Selective Sweeps: An Application to Pooled Sequence Data in Laying Chickens. PLoS ONE 10.1371.

24. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, KarolchikD et al. (2011) The UCSC Genome Browser database: update 2011. Nucleic Acids Research. 39:D876-82.

25. Illumina (2010) Data Sheet: DNA Genotyping. CanineHD BeadChip Retrieved November 15, 2013 from

http://res.illumina.com/documents/products/datasheets/datasheet_caninehd.pdf/.

26. Stothard P, Choi JW, Basu U, Sumner-Thomson JM, Meng Y et al. (2011) Whole genome resequencing of Black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics. 12:559-572.

27. Fan WL, Ng CS, Chen CF, Lu MJ, Chen YH et al. (2012) Genome-wide patterns of genetic variation in two domestic chickens. Genome Biology and Evolution. 5(7):1376-1392.

28. Kim RN, Kim DS, Choi SH, Yoon BH, Kang A et al. (2012) Genome Analysis of the Domestic Dog (Korean Jindo) by Massively Parallel Sequencing. DNA Research. 19(3):275-288.

29. Tsai KL, Starr-Moss AN, Venkataraman GM, Robinson C, Kennedy LJ et al. (2013) Alleles of the major histocompatibility complex play a role in the pathogenesis of pancreatic acinar atrophy in dogs Immunogenetics 65:501–509.

30. Wagner JL, Burnett RC, Storb R (1999) Organization of the canine major histocompatibility complex: current perspectives. J Hered 90:35–38.

31. Huang EJ, Manova K, Packer AI, Sanchez S, Bachvarova RF etal. (1993) The murine steel panda mutation affects kit ligand expression and growth of early ovarian follicles. Developmental Biology. 157(1):100-109.

32. Arkell R, Beddington RS (1997) BMP-7 influences pattern and growth of the developing hindbrain of mouse embryos. Development. 124(1):1-12.

CHAPTER VII

CONCLUSIONS

SUMMARY

Next-generation sequencing (NGS) technology has helped with the construction of *de novo* genome sequences and the mapping of other genomes, using reference sequences (1-6). In addition, NGS has enhanced the ability to gather the following genome-wide sequence information: explaining phenotypes by sequencing genotypes, identifying polymorphisms, mapping mutations, and identifying noncoding RNA (7). Whole-genome sequencing is a thorough way to investigate a genome; however, single nucleotide polymorphisms (SNPs) and BeadChip arrays use information from a previously sequenced genome to interrogate a genome more quickly and affordably by using genome-wide association studies (GWASs).

GWASs have been used successfully within and across breeds to identify loci and mutations underlying diseases, physical characteristics, and behavioral traits of dogs (8–16). When using this approach, two important factors that must be accounted for are population stratification and kinship within the affected and control individuals (17). For the work presented here, we collected unrelated dogs and chickens for our GWASs and whole-genome resequencing (12,18). In addition, we conducted principal components analysis to detect population stratification (15). GWASs have allowed our laboratory to study more quickly diseases and other phenotypic traits in the dog.

OBJECTIVES

The goals of the work presented here were as follows: Chapter II - identify the genetic mutation that causes the autosomal recessive disease episodic falling syndrome (EFS) in the Cavalier King Charles Spaniel. Chapter III - identify loci associated with four diseases (pituitary dwarfism, degenerative myelopathy, congenital megaesophagous (ME), and pancreatic acinar atrophy (PAA)) in the German shepherd dog. Chapter IV - identify the loci associated with the tailless and ear-tuft phenotypes in the Araucana chicken. Chapter V - identify loci associated with dermatomyositis (DM) in the collie. Chapter VI - generate whole-genome resequencing data for the collie in order to annotate single nucleotide polymorphisms (SNPs), insertions, deletions, inversions, tandem duplications, break points, and selective sweeps present in the breed relative to the reference genome.

EPISODIC FALLING SYNDROME IN CAVALIER KING CHARLES SPANIELS

Chapter II reports that a 15.7 kb deletion in the *BCAN* gene causes EFS in the Cavalier King Charles Spaniel (12). Preliminary studies indicated that EFS is inherited in an autosomal recessive fashion, suggesting that relatively few individuals could be used to identify the associated locus. A GWAS, using Affymetrix Canine V2 SNP arrays data from 12 individuals (five affected and seven control) and Sanger sequencing of positional candidate genes, allowed for the discovery of this mutation (12). The 12 study dogs came from the USA, New Zealand, and the UK. Additional testing of unrelated American Cavalier King Charles Spaniels revealed a 12.9% carrier rate, making the mutation more

143

common than expected (12). Finally, the use of the developed direct mutation test will allow breeders to identify carriers and prevent litters with affected puppies (12).

MULTIPLE DISEASES IN THE GERMAN SHEPHERD DOG

A single data set was interrogated for association with both simple and complex diseases in Chapter III. Affymetrix arrays were used to generate SNP profiles for 197 German shepherd dogs, segregating multiple phenotypes. A GWAS for autosomal recessive pituitary dwarfism used 4 affected and 193 normal-sized German shepherd dogs. Association was detected with a locus on chromosome 9, which included a strong candidate gene, *LHX3* (15).

Degenerative myelopathy is inherited in an autosomal recessive pattern with incomplete penetrance. Significantly associated SNPs proximal to the causative gene, *SOD1*, were identified along with other unlinked loci, suggesting the presence of modifying loci (15). Awano et al. (8) identified a missense mutation in *SOD1* in German shepherd dogs and other breeds as the causative mutation.

ME was thought to be a simple recessive trait, but despite a large population (19 affected and 177 control), a predominant association was not detected (15). ME was associated with a 4.7 Mb haplotype of SNPs on chromosome 12 present in all affected German shepherd dog but also in 46% of the control dogs (15). These results support an autosomal recessive pattern with incomplete penetrance.

144

PAA is a complex, autoimmune disorder. GWAS, using 100 affected and 79 controls, showed significant associations throughout the genome, suggesting either the presence of multiple loci with small effects or that PAA may be a heterogeneous disorder (15).

RUMPLESSNESS AND EAR-TUFTS IN ARAUCANA CHICKENS

In Chapter IV, SNP data generated using an Illumina platform was utilized to identify loci associated with two traits of the Araucana chicken. Mapping for the rumpless (*Rp*) and ear-tuft (*Et*) phenotypes in the Araucana chicken revealed associations on chromosome 2 and 15, respectively (18). *Rp* is associated with a 2.14 Mb haplotype, which includes a critical region of 0.74 Mb having two homeobox genes: *Irx1* and *Irx2* (18). *Et* is associated with a 0.58 Mb haplotype in all tufted chickens (18). Because previous inheritance studies indicate that the trait is homozygous lethal, a 60 kb region for which all tufted chickens are heterozygous was identified and harbors the complete coding region of *TBX1*, a plausible candidate gene (18).

Whole-genome resequencing data for the Araucana chicken was generated to investigate the associated regions on chromosomes 2 and 15. Unfortunately, there were multiple sequencing gaps within both regions. These regions have high GC content, requiring special protocols. Future experiments will need to be carried out in order to fill in these gaps and to identify the mutations causing the traits.

DERMATOMYOSITIS IN THE COLLIE

The genetic basis for DM, an inherited autoimmune disease, was investigated in the collie in Chapter V. The pattern of inheritance for DM was previously reported to be autosomal dominant (19). GWAS identified a genetic region of association on canine chromosome 10, and we identified a 10.5 Mb haplotype present in all affected collies (nine heterozygous and 12 homozygous). This region on canine chromosome 10 shares conservation of synteny with human chromosome 12, which harbors approximately 130 genes.

Collies and Shetland sheepdogs share a common ancestor in the early herding dog of Scotland and have been interbred as recently as the 1950s. As a result, both likely share a founding mutation for DM. Using an across-breed approach allows us to take advantage of the differences in homogeneity that exists between the breeds (16). An across-breed study using Shetland sheepdogs will be carried out in the future in order to refine this 10.5 Mb region, and identify positional candidate gene(s).

MAPPING THE COLLIE GENOME

Chapter VI describes the whole-genome resequencing of the genomes of five collies. A 113-fold coverage of the collie genome was produced, and over 9.7 million SNPs were identified; 671,197 structural variants (deletions, insertions, inversions, tandem duplications, and break points) were found in all 5 collies. Additionally, the identification of selective sweeps indicates recent losses in heterozygosity on the

following chromosomes: 1, 5, 8, 10, 12, 15, 24, 29, and 34. Initial analyses identified genes on chromosomes 12 (MHC genes), 15 (*KITLG*), and 24 (*BMP7*) that may underlie traits fixed in the collie. In addition to validating the existing data, further analyses on copy-number variation will be performed. Furthermore, data generated herein will be used to identify candidate causal mutations for canine DM.

IMPACT

In conclusion, this work provides further validation that whole-genome SNP data can be used to identify SNPs and haplotypes associated with phenotypes. NGS technologies provide an affordable means by which to identify additional SNPs and structural variants specific to a population or an individual. When combined, these tools provide a powerful method for the dissection of heritable traits in dog and chicken.

REFERENCES

1. Seabury CM, Dowd SE, Seabury PM, Raudsepp T, Brightsmith DJ et al. (2013) A multi-platform draft *de novo* genome assembly and comparative analysis for the Scarlet Macaw (Ara macao). PLoS ONE 8(5):e62415.

2. Kudirkiene E, Christensen H, Bojesen AM (2013) Draft genome sequence of Gallibacterium anatis bv. Haemolytica 12656-12 liver, an isolate obtained from the liver of a septicemic chicken. Genome Announc Oct 10 1(5): pii e00810-13.

3. Causse M, Desplat N, Pascual L, Le Paslier MC, Sauvage C et al. (2013) Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. BMC Genomics Nov 14 14(1):791.

4. Stothard P, Choi JW, Basu U, Sumner-Thomson JM, Meng Y et al. (2011) Whole genome resequencing of Black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics. 12:559-572.

5. Fan, W-L., Ng, C.S., Chen, C-F., Lu, M.J., Chen, Y-H., Liu, C-J., Wu, S-M, Chen, C-K., Chen, J-J., Mao, C-T., Lai, Y-T., Lo, W-S., Chang, W-H., Li, W-H. (2012) Genome-wide patterns of genetic variation in two domestic chickens. Genome Biology and Evolution. 5(7):1376-1392.

6. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E et al. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. Nature Mar 25 464(7288):587-91.

7. Mardis, E.R. (2008) The impact of next-generation sequencing technology on genetics. Trends in Genetics. 24(3):133-141.

8. Awano T, Johnson GS, Wade CM, Katz ML, Johnson GC et al. (2009) Genome-wide association analysis reveals a SOD1 mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis. Proc Natl Acad Sci USA 106: 2794–2799.

9. Frischknecht M, Niehof-Oellers H, Jagannathan V, Owczarek-Lipska M, Drögemüller C et al. (2013) A COL11A2 mutation in Labrador retrievers with mild disproportionate dwarfism. PLoS ONE 8(3):e60149.

10. Pollinger JP, Bustamante CD, Fledel-Alon A, Schmutz S, Gray MM et al. (2005) Selective sweep mapping of genes with large phenotypic effects. Genome Research 15(12): 1809- 1819.

11. Ostrander EA, Giger U, Lindblad-Toh K (Eds.) (2006) The dog and its genome. Cold Spring Harbor, NY. Cold Spring Harbor Laboratory Press.

12. Gill JL, Tsai KL, Krey C, Noorai RE, Vanbellinghen JF et al (2011) A canine BCAN microdeletion associated with episodic falling syndrome. Neurobiol Dis 45(1):130-136.

13. Hytönen MK, Arumilli M, Lappalainen AK, Kallio H, Snellman M et al. (2012) A Novel GUSB mutation in Brazilian Terriers with severe skeletal abnormalities defines the disease as mucopolysaccharidosis VII. PLoS ONE. 7(7):e40281.

14. Meurs KM, Mauceli E, Lahmers S, Acland GM, White SN et al. (2010) Genome-wide association identifies a deletion in the 3' untranslated region of a Striatin in a canine model of arrhythmogenic right ventricular cardiomyopathy. Human Genetics 128(3):315-324.

15. Tsai KL, Noorai RE, Starr-Moss AN, Quignon P, Rinz CJ et al. (2012) Genome-wide association studies for multiple diseases of German shepherd dogs. Mammalian Genome 23(1-2):203-211.

16. Sutter NB, Ostrander EA (2004) Dog star rising: the canine genetic system. Nat Rev Genet 5:900-910.

17. Hoffman GE (2013) Correcting for population structure and kinship using the linear mixed model: theory and extensions. PLoS ONE 8(10):e75707.

18. Noorai RE, Freese NH, Wright LM, Chapman SC, Clark LA (2012) Genome-wide association mapping and identification of candidate genes for the rumpless and ear-tufted traits of the Araucana chicken. PLoS ONE 7(7): e40974.

19. Hargis AM, Haupt KH, Hegreberg GA, Prieur DJ, Moore MP (1984) Familial canine der- matomyositis: Initial characterization of cutaneous and muscular lesions. The American Journal of Pathology 116: 234-244.

20. Earl D, Bradnam K, St. John J, Darling A, Lin D et al. (2011) Assemblathon 1: A competitive assesement of de novo short read assembly methods. Genome Research: 2224-2241.

21. Illumina (2010) Technical Note: Sequencing. De Novo Assembly Using Illumina Reads Retrieved November 15, 2013 from http://res.illumina.com/documents/products/technotes/technote_denovo_assembly_ecoli.pdf/.

APPENDICES

Supplementary Information for Chapter 2

Gill et al Suppl. Table 1. Primer sequences for canine *BCAN* and *HAPLN2* exon amplification.

Gene	Exon	Forward Primer	Reverse primer
BCAN	1	cttctctccagaaccatgtcctac	acacagcaagtaagtggcagagtt
	2	gagttaacggtgaggttgggacagt	cccacagtgccctctatcctatctc
	3	aagatttggggacagatctggagag	ccccaagagagaaaggaggtaacaa
	4	ctttctctcttggggtggacagact	ggtcctggatggttgacctgag
	5	gggtgtctctgcagaagaaaacaat	gaagacctctggacagcaccg
	6	ggcccagagaagccagccta	gggaaacttcagagctcaagtctgt
	7	ccacagggaagatgagtgagaattg	ccagcatcactctggacacctt
	8	ctcaccctccacagcccctt	ccagagatcatgtgacccagagctt
	9	gtgatagctcccaagacaaggagat	gcagggtccaggcttcaggtcta
	10	ctctgctggctctctggcat	tgagtgggagagagcaggtga
	11	acggacagggaagcagggaa	gcggaggcagaagtgcttgg
	12	tgggcctccactcctcatcc	ttacacacatgcggcttgggtc
	13	atcccgggtctccaggatca	gagcccagggctactgttggata
	14	ccagggctagtgtttggatgagat	ccctcacgtggtcacttcctatg
HAPLN2	1-2	gacattccccacacacaccaag	cgcctgtttcgcttcatagtaattg
	3-5	tgtctgccggctctccctaa	aggatacagacccatcctgaagcac

Gill et al Suppl. Table 2. Chromosome 7 single nucleotide polymorphisms associated with EFS. A total of 17 SNPs on canine chromosome 7 were associated with EFS (P_{raw} values ≤ 0.0001).

SNP	P-value
7.43389066	5.10×10^{-7}
7.46204875	5.61×10^{-6}
7.46283892	5.61×10^{-6}
7.42051505	7.65×10^{-6}
7.39115202	6.12×10^{-5}
7.39142101	6.12×10^{-5}
7.43140878	6.73×10^{-5}
7.43197311	6.73×10^{-5}
7.43230030	6.73×10^{-5}
7.43298329	6.73×10^{-5}
7.43302986	6.73×10^{-5}
7.43867010	6.73×10^{-5}
7.43917092	6.73×10^{-5}
7.44014865	6.73×10^{-5}
7.44059025	6.73×10^{-5}
7.41416845	7.19×10^{-5}
7.42027729	7.19×10^{-5}

Gill et al Suppl. Table 3. Probes for MLPA

Probe	Sequence	Tm°C	%GC	Size
CFTRa	GGGTTCCCTAAGGGTTGGAGCTCCATTGCAATCTACCTAGCCATTG	72	48	100
	P-GCTTATGCCTCTCTTTATCATGAGGCCGCTTCTAGATTGGATCTTGCTGGCAC	75	48	
BCAN PR	GGGTTCCCTAAGGGTTGGAGGATTGGGTGCCCTGTTTGG	72	60	104
	P-CTCGCCACAGGGAGCCTTCTGCGCATTTCCAAGTGGTCACTGTCTAGATTGGATCTTGCTGGCAC	72	65	
BCAN Exon 1	GGGTTCCCTAAGGGTTGGACTGGTCTAGCCTCTAGGAACCGACGCAGAG	70	62	111
	P-GAGGCAGCGGTAGCGTGACAGGCTGGGGAAGAGCAAAAATCACGGAGTCTCTAGATTGGATCTTGCTGGCAC	71	68	
BCAN Exon 2	GGGTTCCCTAAGGGTTGGACTATGCAAGTGTGGCCTTAGCTGATGCCCTG	72	59	115
	P-GAAGGGGACAGCTCAGGTAAGCAGGAGCCCGAGGGGTGTCCTCTAGATTGGATCTTGCTGGCAC	70	59	
BCAN Exon 3	GGGTTCCCTAAGGGTTGGAGCCATCTACCGCTGCGAGGTCCAGCACG	77	64	96
	P-GCATAGATGACAGCAGCGATGCCGTATCTAGATTGGATCTTGCTGGCAC	72	57	
BCAN Exon 4	GGGTTCCCTAAGGGTTGGAGGGGCTATGAACAGTGTGATGCTG	72	54	88
	P-GCTGGCTATCTGACCAGACCGTTCTAGATTGGATCTTGCTGGCAC	71	59	

Key: P- = 5' phosphate group, Blue: 5' PCR primer target, Green: 3' PCR primer target, Red = stuffer

Observed migration: BCAN Ex3 86 bp; BCAN Ex2 94 bp; CFTRa control 98 bp; BCAN PR control 102 bp; BCAN Ex0 113 bp; BCAN Ex1 118 bp; BCAN Ex0 113 bp.
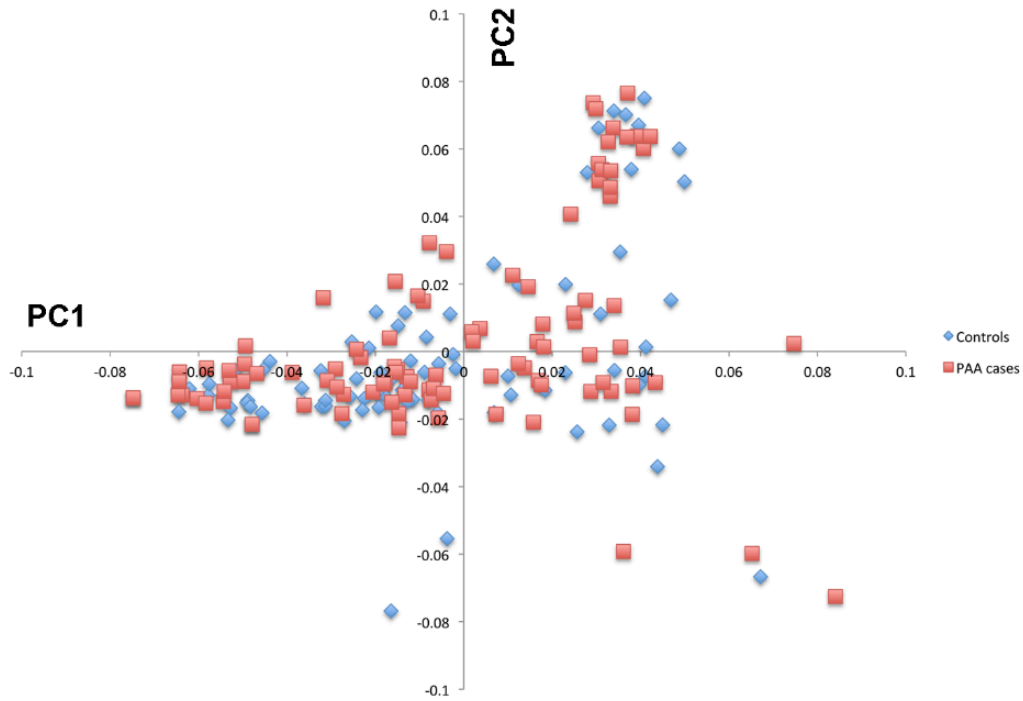
Supplementary Fig. 1 Principal component analysis of 98 GSDs with PAA versus 79 healthy control GSDs. The *x* axis is principal component 1 and the *y* axis is principal component 2. Both populations appear to be evenly distributed throughout the cluster and no significant stratification was observed

Supplementary Table 1: Uncorrected and corrected P values (PAA) are shown for the 50 SNPs with significant P_{raw} values ($P<0.0001$). Corrected P values are from 100,000 permutations (EMP2 or P_{genome}) and the Benjamini-Hochberg test (FDR_BH).

SNP	P_{raw}	EMP2	FDR_BH
chr35.4925078	1.59E-11	2.00E-05	6.43E-07
chr12.58146406	9.28E-11	2.00E-05	1.30E-06
chr28.32676786	9.59E-11	2.00E-05	1.30E-06
chr12.23177472	1.48E-10	2.00E-05	1.50E-06
chr33.30105637	1.85E-10	2.00E-05	1.50E-06
chr15.40294695	3.03E-10	2.00E-05	2.05E-06
chr7.82831339	2.43E-09	0.00026	1.40E-05
chr5.51353689	3.44E-09	0.00035	1.73E-05
chr3.76398656	3.85E-09	0.00039	1.73E-05
chr33.25987767	5.02E-09	0.00043	2.03E-05
chr1.9328042	1.18E-08	0.00083	4.36E-05
chr24.31358922	1.77E-08	0.00115	5.97E-05
chr37.12660891	3.69E-08	0.00225	0.0001149
chr9.10239350	4.07E-08	0.00244	0.0001177
chr6.39598998	2.80E-07	0.01431	0.0007564
chr13.30418286	3.16E-07	0.01574	0.0007722
chr10.13435580	3.24E-07	0.01598	0.0007722
chr35.19780099	3.67E-07	0.01791	0.0008268
chr28.13143663	4.73E-07	0.02195	0.001009
chr4.13107661	7.54E-07	0.03298	0.001527
chr4.4080795	8.72E-07	0.03707	0.001675
chr14.54484895	9.10E-07	0.03814	0.001675
chr23.14905565	9.67E-07	0.04017	0.001703
chr17.44088133	1.47E-06	0.05733	0.002474
chr5.68085129	1.94E-06	0.0711	0.003082
chr2.86073514	1.98E-06	0.07227	0.003082
chr12.12567352	4.83E-06	0.1536	0.007242
chr10.68614890	5.75E-06	0.1737	0.008223
chr3.93468384	5.89E-06	0.1771	0.008223
chr12.27205596	6.76E-06	0.1981	0.009124
chr12.3781476	8.88E-06	0.2478	0.01161
chr7.12305239	1.16E-05	0.2998	0.01469
chr28.38709005	1.41E-05	0.3461	0.01731
chr26.20243802	1.49E-05	0.3599	0.01778

chr12.4698937	1.73E-05	0.3985	0.01999
chr9.54243104	1.86E-05	0.4241	0.02098
chr7.5792722	2.79E-05	0.5349	0.03059
chr2.18072233	3.26E-05	0.5823	0.03474
chr3.29459535	3.47E-05	0.6062	0.03574
chr7.11222056	3.53E-05	0.6125	0.03574
chr6.33280800	4.06E-05	0.655	0.04011
chr12.3845215	4.72E-05	0.7004	0.04552
chr7.12424701	4.92E-05	0.7151	0.046
chrX.87975142	5.00E-05	0.7203	0.046
chr6.18214638	5.42E-05	0.7426	0.04881
chr17.5088861	6.64E-05	0.8006	0.0582
chr7.12310223	6.75E-05	0.808	0.0582
chr31.36637612	8.75E-05	0.8689	0.07383
chr17.52274900	9.12E-05	0.8774	0.07542
chr5.3236960	9.98E-05	0.898	0.08084

Appendix C

Supplementary Information for Chapter 5

Data type	Lib name	No. of reads	Read Length (bp)	Fragment length (bp)	Yield (Mb)	Coverage
Illumina PE	Collie 1	364,699,442	100	478	33,075	-
Illumina PE	Collie 1	371,255,612	100	478	33,606	-
Total	Collie 1	735,955,054	100	478	66,681	24.3
Illumina PE	Collie 2	487,373,242	100	530	38,941	-
Illumina PE	Collie 2	484,341,294	100	530	38,937	-
Total	Collie 2	971,714,536	100	530	77,878	25.1
Illumina PE	Collie 3	264,861,662	100	467	24,184	-
Illumina PE	Collie 3	266,131,024	100	467	24,309	-
Total	Collie 3	530,992,686	100	467	48,493	16.8
Illumina PE	Collie 4	308,226,036	100	526	28,558	-
Illumina PE	Collie 4	308,759,182	100	526	28,613	-
Total	Collie 4	616,985,218	100	526	57,171	21.9
Illumina PE	Collie 5	500,770,100	100	513	37,595	-
Illumina PE	Collie 5	495,968,940	100	513	37,173	-
Total	Collie 5	996,739,040	100	513	74,768	24.9

Supplementary Table. 1 Sequence data from 5 collies, 2 lanes of data per collie.

Appendix D

Permission to Reprint Published Work

Back to view orders

Copy order

Print this page
Print terms & conditions
Print citation information
(What's this?)

Confirmation Number: 11101217
Order Date: 06/20/2013

Customer Information

Customer: Rooksana Norai
Account Number: 3000669046
Organization: Rooksana Norai
Email: rooksan@clemson.edu
Phone: +1 (917)6215308

Search order details by: Choose One   Go

Sort order details by: Order Date   ○ Ascending   ● Descending

Order Details

Neurobiology of disease

Billing Status: N/A

Order detail ID:	63785849
Article Title:	A canine BCAN microdeletion associated with episodic falling syndrome
Author(s):	Gill, Jennifer L. ; et al
DOI:	10.1016/J.NBD.2011.07.014
ISSN:	0969-9961
Publication Type:	Journal
Volume:	45
Issue:	1
Start page:	130
Publisher:	ACADEMIC PRESS

Permission Status: ✓ Granted
Permission type: Republish or display content
Type of use: reuse in a thesis/dissertation
Order License Id: 3173210849866

View details

Note: This item was invoiced separately through our **RightsLink service.** More info

$ 0.00

Mammalian genome : official journal of the International Mammalian Genome Society

Billing Status: N/A

Order detail ID:	63785867
Article Title:	Genome-wide association studies for multiple diseases of the German Shepherd Dog
Author(s):	Tsai, Kate L. ; et al
DOI:	10.1007/S00335-011-9376-9
Date:	Jan 01, 2012
ISSN:	0938-8990
Publication Type:	Journal
Volume:	23
Issue:	1-2
Start page:	203
Publisher:	SPRINGER NEW YORK LLC
Author/Editor:	INTERNATIONAL MAMMALIAN GENOME SOCIETY

Permission Status: ✓ Granted
Permission type: Republish or display content
Type of use: use in a thesis/dissertation
Order License Id: 3173210856764

View details

Note: This item was invoiced separately through our **RightsLink service.** More info

$ 0.00

Total order items: 2   **Order Total: $0.00**

159

Appendix E

Original Cover Art by Alessio Mancino

**DISCOVERY OF
VARIANTS UNDERLYING
SIMPLE AND COMPLEX
TRAITS THROUGH GENOME-
WIDE ASSOCIATION STUDY
AND WHOLE-GENOME
RESEQUENCING**

**A Thesis
Presented to
the Graduate School of
Clemson University**

**In Partial Fulfillment of
the Requirements for the Degree
Doctor of Philosophy
Genetics**

**by
Rooksana Elizabeth Noorai
May 2014**

*Aee y:
ei Ae
a, oiee
ar
ua
aa
Ae eu
ei E.
uy
iae
eo*

*Accepted by:
Dr. Leigh Anne
Clark, Committee
Chair
Dr. Susan C.
Chapman
Dr. F. Alex Feltus
Dr. Keith E.
Murphy
Dr. Michael G.
Sehorn*

*Ccptd b:
Dr. Lgh nn
Clrk, Cmmtt
Chr
Dr. Ssn C.
Chpmn
Dr. F. lx Flts
Dr. Kth.
Mrph
Dr. Mchl G.
Shrn*

A.MANCINO PINXIT MMXIII