ABSTRACT

AGLER, CARYLINE SALZMANN. Investigations of Canine Hereditary Diseases. (Under the direction of Dr. Natasha J Olby).

Canine atopic dermatitis is a heritable, pruritic skin condition that causes erythematous lesions on the face, paws, ears, and ventrum. The West Highland White Terrier (WHWT) breed has a high prevalence of the disease, between 15.4 and 18.7%, making it an ideal breed in which to study the genetics of the disease. Canine atopic dermatitis in the WHWT was investigated using linkage and genome-wide association studies (GWAS). Normal dogs and those affected with atopic dermatitis were recruited through referral from veterinary dermatologists and breed clubs. Pedigrees, health questionnaires, and DNA were collected from dogs fitting the diagnostic criteria. Three families of WHWT and a maximum of 90 dogs were genotyped at 256 microsatellite markers from the Minimal Screening Set 2 across the thirty-eight canine autosomes. Linkage analysis was unable to identify a locus for atopic dermatitis in these dogs; however, the analysis did not include the sex chromosomes and homozygous markers resulted in gaps in genome coverage. Simulations performed using the pedigree structure of the genotyped dogs revealed the inability to detect a mutation of low penetrance, which could explain the lack of detection of a disease locus.

The 58 dogs from the linkage study and 128 additional dogs were genotyped on the Illumina CanineHD BeadChip. A GWAS was performed using 96 case and 87 control WHWT and revealed SNPs associated with the disease; however, none passed a stringent Bonferroni correction for multiple testing. When considering the top associated SNPs, none replicated previous genome-wide association studies performed in WHWT or multiple breeds; however, the candidate genes *HMGCR* and *F2RL1* were present in a region of association. The high prevalence of atopic dermatitis in the WHWT breed and the complex nature of the disease makes it difficult to identify normal individuals as well as to be 100% sure of a normal phenotype.

Hereditary ataxia is a neurodegenerative disease that affects the cerebellum and causes the deterioration of the Purkinje neurons of the cerebellar cortex. Clinical signs of the disease include an uncoordinated gait, hypermetria, intention tremor, and wide-based stance. The disease has been described in the Old English Sheepdog (OES) breed as a slowly progressing ataxia with an age of onset between 6 months and 3.5 years and autosomal recessive inheritance. The disease was investigated in the breed using a GWAS and targeted next generation sequencing. OES DNA samples and pedigrees were collected from dogs and the diagnosis of hereditary ataxia was made based on compatible clinical signs, video of gait, MRI, and necropsy; and varied by dog. Fifty-five dogs were genotyped on Illumina SNP20 and CanineHD BeadChips and a genome-wide association analysis was performed in fifty-four dogs with PLINK and FBAT software using the 12,986 SNPs common to both SNP chips. A region of significant association on chromosome 4 was identified; however, numerous genes with no clear candidates for hereditary ataxia were present in the region. Targeted capture using a Nimblegen array and next generation sequencing using Illumina technology was performed in six OES and covered a continuous 12 Mb including the region of multiple association signals from the GWAS. The successful target capture and sequencing produced 47X coverage. Using the Genome Analysis toolkit, six novel non-synonymous sequence variants which segregated with the disease and expected inheritance in the dogs were identified. Genotyping these variants in additional dogs and additional breeds identified an A to C base change in the first exon of the gene *RAB24* which showed perfect segregation in a larger cohort of OES and Gordon Setters also suffering from hereditary ataxia. The identified SNP is highly associated with hereditary ataxia in the OES and Gordon Setter breeds.

Investigations of Canine Hereditary Diseases

by
Caryline Salzmann Agler

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Comparative Biomedical Sciences

Raleigh, North Carolina

2013

APPROVED BY:

Natasha Olby Committee Chair	Thierry Olivry Committee member
Trudy Mackay Committee member	Alison Motsinger-Reif Committee member

BIOGRAPHY

Caryline (Cary) spent half of her undergraduate career at Northern Illinois University in Dekalb, IL and the other half at the University of North Carolina at Chapel Hill where she received her Bachelor of Science degree in Biology.

ACKNOWLEDGMENTS

There are numerous people with whom this work would not be possible: I could not have asked for a better advisor, Dr. Natasha Olby, to guide me through this difficult but rewarding doctoral process. She is a model advisor, researcher, clinician, and person that I am truly lucky to have worked with. The rest of my thesis committee members Drs. Thierry Olivry, Alison Motsinger-Reif and Trudy Mackay must be acknowledged for their continuous guidance, advice, support, and knowledge. Members of the laboratory also under the direction of Dr. Olby including: Drs. Ganokon Urkasemein and Ji-hey Lim, Tonya Harris, and Kimberly Williams were always willing to lend a helping hand. Dr. Dahlia Nielsen, whose statistical and genetics knowledge and keen insights were instrumental to this work. The researchers at the NIH Neurogenetics laboratory in Bethesda, MA for their genotyping guidance including: Dr. Andrew Singleton, Dr. Dena Hernandez, and Sampath Arepalli. The scientists at the Broad Institute in Cambridge, MA for their next generation sequencing expertise include Drs. Kerstin Lindblad-Toh, Snaevar Sigurdsson, Ruqi Tang, and Noriko Tonamura-who was also a gracious host during my visit to the laboratory. Chris Smith and Ryan Agler for their programming and computer support. I must acknowledge the Westie Foundation of America, Old English Sheepdog Club of America, TarTan Gordon Setter Club, and the AKC Health Foundation for funding this work. Last, but not least, all the dog owners and breeders whose commitment to research made all of this work possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Background and Chapter Aims

Dog as a model organism

In addition to being man's best friend, the purebred dog has become a model system for discovering genetic variants causing human disease. The canine species is made up of over 300 breeds of diverse morphological and behavioral characteristics which have been developed over a relatively short period of time [1]. The inherent structure of modern dog breeds, each breed representing an isolated breeding population with extensive inbreeding and population bottlenecks, has led to reduced genetic heterogeneity, numerous genetic diseases in the species [2], and can enhance genetic risk factors making them more detectable by genetic studies [3]. The large litter size, short generation time, and well-documented pedigrees are features of modern dog breeds which are beneficial to genetic mapping studies. When compared to the mouse model system, the dog also has many advantages. First, canine genetic diseases are naturally occurring whereas in the mouse, disease causing mutations are often induced [4]. Many of these genetic diseases have a corresponding human disease with similar physiology, presentation and response to treatment [4]. Second, dogs are more similar in size and have a more similar lifespan to humans than mice and usually share the same environment as their human companions. Finally, dogs and humans are more genetically similar to one another than humans are to mice, as humans have approximately 650 Mb more genomic sequence that aligns to the dog genome than to the mouse genome [5].

Genetic disease mapping

Identification of Genetic Disease

The mapping of genetic diseases begins with the identification of a disease of interest. Before expending the time and resources in investigating the disease of interest it is important to know whether there is a genetic component to it and the strength of that component. Evidence for a genetic influence can be suggested by a disease being present in families via epidemiological evidence such as; a higher proportion of cases than controls have a relative with the condition, a higher proportion of healthy individuals from a cohort with a family history of the disease develop the disease over time than comparative individuals without a family history, and the correlation of values associated with the disease in related pairs is significantly greater than zero [6]. If a disease is determined to be familial, it does not necessarily mean genes are influencing the susceptibility as the disease could result from infectious or environmental factors, behavior, or similar life styles in related individuals. A statistically significant association between family history and disease can be tested for using a standard epidemiological 2x2 table and calculating an odds ratio (OR) with $OR = (a \times d)/(b \times c)$, or by calculating an attributable fraction due to a positive family history which gives information on the amount of the disease attributable to genes [7]. A prospective measure, a relative risk ratio, can be calculated from the prevalence of disease in relatives of an affected individual to the prevalence of disease in the population with the larger the ratio (>1), the greater the familial aggregation [8]. Heritability is a measurement of how much phenotypic variation in a trait is attributable to genetic variation and can be defined in a broad and narrow sense. In the broad sense definition of heritability (H^2); $H^2 = \frac{V_G}{V_P}$, where (V_G) is the total genetic variance of a phenotype, and (V_P) is the total phenotypic variance [9]. The variance factors contributing to the estimate of heritability can be calculated from data collected on the observed and expected similarity between relatives. In the narrow sense, heritability is the additive genetic variance of the phenotype. Heritability varies between 0 and 1, with 0 indicating the variation of trait is not caused by genetic factors and a heritability near 1 indicating the variation of a trait is largely explained by genetic factors. An estimate of narrow sense heritability can be achieved by using the correlation coefficient obtained by plotting values associated with the trait in offspring and their parents. A heritability of 0.3 is considered useful for genetic analysis [7]. In humans, the study of identical (100%

genetic identity) versus fraternal twins (50% genetic identity) provides a unique mechanism to substantiate a genetic component in that twins usually share the same environment and cultural behaviors or life styles (at least while young). It is expected that affected status should be concordant (identical or monozygotic) or approximately 50% concordant depending on the inheritance pattern (fraternal or dizygotic) if the disease is genetic. Correlation between types of twins can also be used to calculate heredity using the equation; $H^2 = 2(r_{MZ} - r_{DZ})$. Diseases of low heritability will likely need a large sample size to be able to identify genetic variants [10]. Adoptive studies can also be useful to determine genetic versus environmental contributions to a disease provided information can be gathered on both the biological and adoptive parents. Similar to humans, a genetic disease is suspected in the canine species when a higher than expected proportion of individuals within a breed, line, or family is noted with the disease without an environmental cause. To investigate the plausibility of genetic involvement, pedigrees can be examined for the dissemination of the disease. As most breeds have multiple offspring per litter, a short generation time, and well documented pedigrees, they provide a useful mechanism for estimating the heritability of a disease.

Mode of Inheritance

Determining the inheritance model of a genetic disease is important in mapping diseases. Diseases of Mendelian inheritance are of a single locus whereas a complex disease is one which results from multiple genetic loci. Segregation analysis and the estimation of familial relative risks can be used to determine the most likely genetic model. Segregation analysis is a modeling tool used to determine if the disease inheritance pattern in a pedigree is consistent with traditional genetic inheritance models such as autosomal dominant, autosomal recessive, or X-linked, or is more consistent with polygenic or environmental models. Different methods of segregation analysis have been developed to accommodate modes of family ascertainment, how the family came to be included in the study; be it through identification of an affected offspring or via random sampling of parents. Simple segregation analysis methods such as the Davie's extension of the singles method and the direct method of maximum likelihood estimation or complex segregation analysis methods based on the Elston-Stewart algorithm can be used [1, 11]. The mode of inheritance of a disease in canine pedigrees can also be investigated with simple segregation or complex segregation analysis. Many of the genetic diseases identified in dog breeds have been assigned a specific mode of inheritance; however, often this is done using a pedigree analysis only as an insufficient sample size has not enabled statistical evaluation [12]. Planned matings can also be undertaken to test a hypothetical mode of inheritance.

Study Design

With a disease identified and genetic component confirmed with a consistent genetic model, the study design is determined. Samples must be collected which may be of a single affected individual, relative pairs, or an extended family with multiple affected individuals [7]. An extended pedigree with many affected individuals and multiple generations is typically used in a linkage study; however, association studies can also be accomplished with this ascertainment method. When control samples are collected, it is important that they come from the same population and are as similarly matched to case samples as possible. Ascertainment bias can influence the ability to detect a true variant and may lead to false positive results.

An important aspect in performing any genetic mapping study is the power of the study or its ability to detect a causal variant or, statistically speaking, the probability of rejecting the null hypothesis. In cases of simple Mendelian inheritance, power studies are fairly straightforward and can be assessed by simulations using computer programs. To assess power in linkage studies, information on the genetic model, penetrance, frequency of the trait allele, information on the genetic markers as well as the pedigree and phenotypes need to be included in the simulation process and the results inform upon the ability to detect linkage, the type I error, or α (the rate of incorrectly rejecting the null hypothesis), and type II error, or β (the rate of incorrectly rejecting the alternative hypothesis), rates [7]. In case-control studies, known parameters for assessing power include sample size, type I error, and disease prevalence as well as unknown parameters such as the mode of inheritance, effect size, marker allele frequencies, and extent of linkage disequilibrium [13]. Formulae for determining power without performing simulations are available for the transmission disequilibrium test and case-control tests for association as well as online tools such as the Genetic Power Calculator are available [14-16]. Some of the parameters for estimating power such as marker allele frequencies and linkage disequilibrium have been examined in multiple human populations (Caucasian, Japanese, Chinese, and African). This would be difficult to do in dog breed populations as there are over 300 breeds; nevertheless, sample size estimates have been made to provide the sufficient power to detect variants causing a risk for a disease. Coalescence modeling was used to determine the number of SNPs and individuals needed to perform genome-wide association studies in a breed. Approximately 15,000 SNPs provide sufficient power and the use of 20 cases and 20 controls would allow detection of a simple Mendelian highly penetrant recessive trait with no phenocopies (non-genetic causes for the phenotype) and 50 cases and controls would be needed for a dominant trait [1, 17]. This number of SNPs would not be sufficient in performing across-breed studies and may not be sufficient in breeds such as the Labrador Retriever as the breed has shorter levels of LD and larger haplotype diversity [1]. While 20 cases and 20 controls have been proposed, fewer have successfully been used to map recessive diseases; 12 cases and controls for generalized progressive retinal atrophy in the Irish Glen of Imaal Terrier [18] and 13 cases/21 controls for exfoliative cutaneous lupus erythematosus in German Shorthaired Pointers [19]. Using a simple multiplicative model for a trait conferring a 5-fold increased risk, a sample size of 100 cases and 100 controls would give 98% power and decrease to 50% if the trait conferred a 2-fold risk [17]. Fewer than 100 cases and 100 controls in complex disease mapping has been successful as 81 cases and 57 control Nova Scotia Duck

Trolling Retrievers were used to identify three genome-wide significant loci for a SLE-related disease complex [20]. Genetic marker data from a specific dog breed which gives information on the allele frequencies and LD patterns can be used to estimate power in a mapping study of a related dog breed.

An important aspect of the study design that must be considered is what portion of the genome will be assessed and how. Genetic variation at a locus is termed a polymorphism if it occurs with a frequency of at least 1% in the population. A candidate gene approach examines polymorphisms in genes suspected to play a role in the disease of interest; whereas, a genome-wide approach examines polymorphisms spanning the entire genome. Homozygosity or autozygosity mapping can be used to map recessive diseases, as affected individuals are expected to be homozygous in the genomic region harboring the causative mutation.

Genetic Markers

The mapping of genetic diseases requires genetic markers which are based on DNA polymorphisms. The first genetic markers used for mapping were restriction fragment length polymorphisms (RFLP) of which a polymorphism in the DNA sequence causes a restriction enzyme to cut the DNA where it previously did not, or to no longer cut a site where it was previously able to. Polymorphisms composed of multiple nucleotides include the variable number of tandem repeat (VNTR) or minisatellites, and microsatellites; also known as simple sequence repeats (SSR) or short tandem repeats (STR). Microsatellite markers are regions of the genome where nucleotides (typically two, three, or four) are repeated in tandem. They have been widely used in linkage studies as they have a high heterozygosity, are a relatively inexpensive technology with high throughput capability, and are easy to decipher alleles and perform quality control on [21]. Single nucleotide polymorphisms (SNP) are single base changes in the DNA sequence, are the most common source of genetic variation, and are the most common type of genetic marker used in association studies. The canine genome consists of 40 chromosomes which includes 38 pairs of acrocentric autosomes and the 2 metacentric sex chromosomes, X and Y [22] in which numerous genetic markers have been characterized. The initial canine linkage map was based on polymorphic microsatellite markers. The map is created by linkage mapping in which the genotyping of markers in individuals of multi-generational pedigrees is performed followed by enumeration of the recombinant and non-recombinant offspring, allowing the calculation of the recombination fraction between markers. The recombination fraction (θ) can then be converted to a physical distance to compose the map. The Kosambi function is one conversion of recombination fraction into map distance and is $x = \frac{1}{2}\tan^{-1}(2\theta) = \frac{1}{4} Ln\frac{1+2\theta}{1-2\theta}$; where (x) is the map distance in Morgans [12]. One percent recombination is approximately equal to 1 centiMorgan. The canine linkage map has morphed from the initial 139 microsatellite markers compromising 30 linkage groups [23] to the current map of approximately 3000 microsatellite markers specifically assigned to each of the canine chromosomes [24]. In parallel with the expansion of the canine linkage map, microsatellite markers organized into sets for ease of whole-genome genotyping have been developed and described. The Minimal Screening Set 3 contains 507 markers with the density of a marker approximately every 5 centiMorgans [25]. Additional resources which have been instrumental in expanding the knowledge of the canine genome include a radiation hybrid (RH) map which incorporates 4249 markers including genes, microsatellites, and bacterial artificial chromosomes (BAC) clones [26]. A subsequent RH map used the 1.5X survey sequence of a Standard Poodle [27] to place over 10,000 gene markers [28]. The ultimate genetic tool has become available fairly recently; the high quality 7.5X genome sequence of a Boxer and 100,000 random sequence reads from nine unrelated dog breeds [5]. This has enabled the identification of over 2.5 million SNPs as well as allowed the refinement of the canine linkage and RH map [24, 25, 29, 30]. Of the 2.5 million SNPs identified in 12 dog breeds, approximately 72% of the SNPs are polymorphic in a specific breed and multiple arrays of SNPs have been developed for genotyping in all dog breeds [1]. While versions of canine whole-genome genotyping SNP chips from Affymetrix (27K and 50K) and Illumina

(22,362 SNPs) have been developed, the current Illumina CanineHD chip offers an increased SNP density and contains over 173,000 SNPs with a uniform genetic coverage which can improve the chance of identifying a disease risk allele in particular when using haplotype-based and homozygosity mapping [31]. This chip was designed after additional sequencing of four dog breeds and has a mean SNP spacing of 13 kb and improved coverage of the X chromosome [1].

Linkage Analysis

One method used for gene mapping is linkage analysis. Linkage aims to identify regions of the genome where a trait is co-inherited with a genetic marker using information from pedigrees. Linkage mapping is based on the measure of recombination or crossing-over of chromosome segments taking place during meiosis and relies on the fact that when two positions on a chromosome are closer together less recombination takes place between the two; unlinked loci have a recombination fraction of 0.5 and segregate independently [12]. Two loci are linked if the recombination fraction between the two deviates significantly from 0.5. Linkage analysis is generally categorized into model-based (parametric), or when performed without a genetic model, non-parametric.

Parametric linkage analysis is a model-based analysis method in which the data is examined to see if they fit a particular genetic model and includes specification of the mode of inheritance of the trait, the trait and marker allele frequencies, and the penetrance of the disease genotype. The penetrance of a trait reflects whether an individual with the specific genotype will exhibit the phenotype. Incomplete penetrance could indicate gene-gene interactions or environmental factors. The data are tested to determine whether the recombination fraction (a value between 0 and 0.5), between two loci is significantly different from 0.5 [32]. The most common method of parametric linkage analysis is the LOD score approach [33]. In this approach, a logarithm of odds (LOD) ratio is used as a statistical measure to support or reject evidence for linkage. The LOD score is defined as; $Z(\theta) = log_{10}(\frac{L(\theta)}{L(\theta=.5)})$, where ($Z(\theta)$) is the LOD score, ($\theta$) is the recombination fraction, and (L) is the likelihood. A LOD score of 3.0 corresponding to a 1000 to 1 odds ratio is generally considered significant threshold for linkage with a LOD score less than -2.0 a threshold against linkage [33]. In two-point LOD score analysis a single marker is examined in relation to a trait locus; whereas, in multi-point analysis examination of the trait locus is in relation to a fixed set of markers of a known distance. In multi-point linkage analysis the inter-marker distances along with the other model parameters need to be specified. In genome-wide linkage scans, a genome-wide significance LOD score threshold of 3.3 corresponding to an alpha of 0.05 has been proposed [34]. Some of the algorithms implemented for use in multi-point linkage analysis include the Elston-Stewart [35], Lander-Green [36], and Markov chain Monte Carlo (MCMC). Linkage studies are particularly amenable to searching for Mendelian traits, over 200 have been characterized in the dog [2], representing most disease genes that have been identified thus far. The large multi-generational pedigrees of individual dog breeds are beneficial for mapping using linkage studies; however, the typical canine pedigree structure with inbreeding loops can prove challenging for some software programs which require the loops to be broken. The lm_bayes program in Morgan uses a MCMC algorithm and can handle these pedigree loops as well as the large number of marker data generated from multi-generational pedigrees [37]. Linkage analysis was used in the mapping of diseases such as copper toxicosis in Bedlington Terriers [38], progressive rod-cone degeneration [39], early retinal degeneration [40], a renal cancer syndrome [41], neuronal ceroid lipofuscinosis in English Setters [42], and narcolepsy in Doberman Pinschers [43] with gene localization aided by known syntenic regions of the human genome. Using the pedigree structure used in two canine disease studies for autosomal traits, data was simulated and analyzed to calculate analytic and empirical LOD score thresholds which were found to be 3.2 and 2.7 [44].

Non-parametric or model-free linkage analysis tests whether inheritance deviates from what is expected under independent assortment [45]. This analysis method is more suited to the study of complex traits as it does not need a mode of inheritance specified; however, other parameters are necessary [46]. The methods for model-free linkage analysis examine the sharing of alleles between relatives as it is expected that affected relatives will share an excess of alleles in a region containing a disease causing gene [47]. Shared alleles can be identical by state (IBS) or identical by descent (IBD). Alleles that are IBS are those that look the same but derive from different ancestral sources, and alleles that are IBD are not only the same but are copies of an ancestral allele [32]. Based on the random segregation of gametes during meiosis and the number of meioses separating relative pairs, the expected proportion of the number of alleles IBD and the proportions estimated in marker data from relatives can be compared. The affected sib-pair (ASP) method of linkage analysis compares siblings' proportions of alleles IBD in the marker data for deviation from the expected proportion of IBD alleles under the assumption of no linkage; with the proportions of sharing zero, one, or two alleles being 0.25, 0.5, and 0.25 respectively. Statistical tests include a chi-square goodness of fit for testing deviation of the three proportions from the expected; a one-sided test for deviation from the proportion of two alleles IBD, and a one-sided means test in which the mean number of alleles IBD is compared with the expected value of 0.5 [7]. The difference in IBS alleles in affected sibling pairs can also be examined using a chi-square goodness of fit test. Methods have been developed which examine affected relative pairs; one such method, the affected pedigree member (APM) uses the deviation of IBS alleles and the genotypes of affected individuals, the marker allele frequencies, and the relationships of these individuals [7, 48]. Another affected relative pair method, NPL (nonparametric linkage) was designed for multi-point linkage analysis and can compare all affected relatives simultaneously [45]. Software available for performing non-parametric linkage analysis includes Genehunter [45] and Merlin [49]. While non-parametric linkage analysis is not as commonly used in dogs it has been used with microsatellite markers in candidate genes for the investigation of linkage to cataracts in the wire-haired Dachshund and patellar luxation in Pomeranian dogs [50, 51].

Association Analysis

A second mapping strategy, association analysis, looks for a significantly increased or decreased frequency of a marker allele, genotype, or haplotype with a trait than would occur by chance with no association. An association with a trait can be detected due to the direct biological action of the marker tested or due to a marker being in linkage disequilibrium with the causative locus. Linkage disequilibrium (LD) is the non-random association of alleles at different loci in a population. There are multiple measures of LD which depend on the allele frequencies of two loci ($\mathrm{Pr}(A)$ and $\mathrm{Pr}(B)$) and haplotype frequencies ($\mathrm{Pr}(AB)$) and include the disequilibrium coefficient (D); $D = \mathrm{Pr}(AB) - \mathrm{Pr}(A)\,\mathrm{Pr}(B)$, correlation coefficient (r^2); $r^2 = D/\{\mathrm{Pr}(A)\,[1 - \mathrm{Pr}(A)]\,\mathrm{Pr}(B)\,[1 - \mathrm{Pr}(B)]\}$, and Lewontin's ($D'$); defined as the ratio of D to it maximum possible absolute value given the allele frequencies [52]. A haplotype frequency is the frequency with which two alleles occur in the same gamete. These three measures of LD are zero if there is no allelic association. The sequencing of the canine genome has enabled the comparison of LD and haplotype structure in select dog breeds [5]. The relatively large within breed LD spans kilobases to megabases and is approximately 100 times larger than the human genome with haplotype blocks of up to several megabases interspersed with blocks of high heterozygosity [1]. In dogs as a species, the interbreed LD appears to be shorter than in humans and haplotype blocks are shorter and are shared between distantly related breeds [1, 17]. While LD in dog breeds has been examined in five breeds across 1 Mb regions [53] and in ten breeds across 15 Mb regions [5], a more recent study surveyed LD in 18 breeds across 5.2 Mb of sequence on chromosome 1 and in SNPs in four additional chromosomes [54]. The distance at which the LD correlation coefficient measure decays to 0.2 using SNP data was determined to vary between breeds and ranged from 20 kb in the Labrador Retriever to 4.2 Mb in the Pekingese [54]. The relatively large within breed LD which spans kilobases to megabases suggests that genetic mapping can successfully be accomplished with fewer markers and samples than is necessary in human studies [5, 53, 55]. The analysis of the canine genome has also allowed the examination of the feasibility and requirements of genome-wide association studies in dog breeds. While association studies are geared towards complex disease/traits in human studies, they have been used to identify both simple and complex disease/traits in dogs. The pattern of large haplotype blocks within a specific breed and much smaller haplotype blocks across multiple breeds has led to the proposal of a two-stage mapping strategy of first localizing a trait to a relatively large region in a single breed and subsequent fine mapping across different breeds provided the same mutation responsible for the trait is present in multiple breeds. The genome-wide genotyping of dogs using SNPs has become incredibly affordable and has precipitated numerous studies of canine genetic diseases. In addition, the rapid advancement of sequencing technologies and great reduction in cost of sequencing makes it possible to sequence the entire genome of individuals [56].

In population based case-control association studies, SNPs are the commonly used genetic markers and can be analyzed using a contingency table or logistic regression. Association can be tested for by comparing the frequencies of each allele of the SNP or the genotype classes against the trait to determine whether there is a difference in frequencies between cases and controls using a simple chi-squared test for independence of rows and columns of the contingency table [57]. In addition to counting the genotypes, different genotypic models such as recessive, dominant, multiplicative and additive can be considered and tested [58]. The dominant model considers having one or more copies of an allele to increase risk for the trait; whereas, in the recessive model, having two copies of an allele increases risk [58]. In multiplicative and additive genetic models, the risk for the trait increases by a factor in heterozygotes with that factor doubled in the additive model and squared for homozygotes in the multiplicative model [59]. The Cochran-Armitage test for trend can also be used to test for difference in genotypes as it is more robust against deviation from Hardy-Weinberg equilibrium (when the genotype frequencies in the population are dependent only on allele frequencies) [32, 60]. Logistic regression can be used in analysis of a single marker, multiple markers, and when additional covariates need to be incorporated. In the analysis of quantitative traits, traits assumed to be influenced by many loci contributing a small effect, linear regression and analysis of variance (ANOVA) are statistical methods commonly used [61]. One method used for investigating quantitative trait loci (QTL) involves crosses of distinct inbred lines for the trait of interest in which each inbred line is assumed to be fixed for opposite alleles influencing the trait of interest and breeding of the two lines and the backcrossed generations are examined at polymorphic genetic markers for segregation with the phenotype [9].

Ideally in population based case-control association studies, cases and controls will be ascertained from the same population; however, this does not always happen and it can be difficult to recognize when they are not. Whether samples are coming from different populations should be considered when performing case-control association studies as it can lead to false positive associations due to differences in the trait and allele frequencies of subpopulations. Population structure or population stratification: sample structure due to differences in genetic ancestry and therefore individuals being sampled from populations with different ancestries [62], cryptic relatedness: individuals being closely related and therefore not independent samples [63], or family structure can all lead to spurious associations. Numerous methods to correct for sample structure have been developed and are implemented in programs such as STRUCTURE [64], EIGENSTRAT [65, 66], TASSEL [67, 68], and EMMAX [62]. Genomic control is one way to account for population substructure by calculating the test statistics at SNP loci unlikely to have an effect on the trait and determining an inflation factor then using that inflation factor to adjust the test statistics [69, 70]. An inflation factor greater than 1 is evidence for stratification or other confounding factors [57]. A structured association method uses genotype correlations between unlinked markers from a sample of individuals to assign them to separate subpopulations then tests for association within these subpopulations [71]. Principal components analysis in which principal components are identified and used as covariates in analysis of the genetic data has also been used to correct for population structure [65]. Mixed model approaches using kinship matrices inferred from the genotype data can be effective in correcting for population structure, genetic relatedness, and family structure seen in model organisms by modeling both fixed and random effects [72]. Population structure occurs in dog breeds and can be assessed via pedigree analysis or genotyping data and the techniques described above can be used to detect and correct for it when present [73].

A family-based design for association analysis can be used instead of a case-control design. The family-based design is robust against bias caused by population stratification [74]. Tests for association using a family triad consisting of an affected offspring and both parents include the haplotype relative risk (HRR), AFBAC, and the transmission/disequilibrium test (TDT) [7]. The transmission disequilibrium test (TDT) surveys for a difference in allele transmission from a heterozygous parent and their affected offspring and tests for both linkage and association between the marker and phenotype [16]. The family-based association tests (FBATs) are an extension of the TDT method with one such extension being implemented in the FBAT toolkit [75]. This toolkit improves upon the TDT method by handling factors such as: missing parents, additional family members, and different genetic models and phenotypes [75]. As diseases in dog breeds are often seen in pedigrees and often many family members can be easily genotyped, family-based association methods are useful in avoiding spurious associations due to population structure.

In addition to a single marker being tested for association with a trait, multiple markers making up a haplotype can also be tested. A haplotype is the combination of alleles on a chromosome or portion of a chromosome that are inherited together. The trait is tested for association with the marker haplotypes. As genotypes are collected using genetic markers and not the haplotypes, the phase, which nucleotides are present together on a particular chromosome, and therefore haplotypes must be inferred. Inference of haplotypes can be done with statistical algorithms such as the expectation-maximization (EM) algorithm [76], Clark's parsimony algorithm [77], and those using Bayesian models [78, 79]. Haploview, a tool for analysis and visualization of LD and haplotype maps, can be used to perform association tests on haplotypes as well as single markers [80, 81] as can FBAT for family-based data.

Association studies usually test numerous markers for association with a trait; therefore a correction for multiple testing should be done to account for the increased likelihood of a false positive result that can occur with so many individual tests. A Bonferroni correction does this by adjusting the type I error rate, by dividing it by the number of tests performed. However, a Bonferroni correction is conservative because it assumes that each test is independent when in fact some SNPs are correlated due to LD. A false discovery rate (FDR) procedure which controls for the expected proportion of falsely rejected hypotheses can also be used to adjust for multiple testing [82]. Permutation testing is a common method used to correct for multiple testing. In permutation testing, the case and control status of the samples are reshuffled and association tests performed, the final marker p-values are adjusted based on the number of times the permuted data has a test statistic greater than the original dataset [83]. With the extensive LD in dog breeds, a Bonferroni correction can be extremely conservative; therefore, permutation testing is the more common method of choice for multiple testing correction in genome-wide association studies.

Follow-up

A significant finding in a genetic mapping study is usually just the beginning of the arduous process of determining the causal variant or functional effect of the identified result. When linkage mapping with microsatellites, the significant locus identified will likely be large and span megabases. On the other hand, mapping using SNPs will likely result in a smaller identified locus indirectly correlated with the disease due to being in LD with it. Complex diseases may be influenced by genetic heterogeneity, multiple interacting genes, known as epistasis, or gene-environmental effects. Genetic heterogeneity occurs when several genes cause the same disease as opposed to allelic heterogeneity, where different mutations within a gene cause the same disease. Epistasis may result from a single modifier gene or many interacting genes. Approaches to identifying these gene interactions or a gene-environment interaction may include sophisticated statistical modeling and more than one of the analysis methods described previously [7]. Replication of association results is an important step in validating associations for a disease; yet, is one of the most challenging. Ideally, a replication study would show a significant association in an independent sample with a similar effect size at the same allele for the disease as the initial study. Two strategies for replication studies include an exact strategy where only positively associated markers are genotyped and local strategy where the positively associated markers along with additional markers are genotyped [57]. A multistage study design aimed to replicate association findings using a single study is devised with an initial stage of scanning the genome for associated loci by genotyping a portion of the total sample size, and a follow-up stage to further investigate the promising signals found in the initial stage by genotyping the remaining collected individuals [84]. Collaborations between researchers allowing meta-analysis of data can increase sample sizes and be beneficial, especially when mapping complex diseases. In addition to the replication of initial findings, functional data which investigates the molecular change and resulting biological process is important to show a variant is causative for the disease. Some of the methods for this include *in silico*, *in vitro*, and genetic manipulation of model systems [84]. The identification of disease causing genes relies on the ability to determine the location of genes in the genome. While the Canfam 2 genome assembly of the dog is of high quality, it does have some large unreliable regions and small gaps making gene annotations incomplete. The Canfam 3 version of the genome has filled approximately half the gaps and continued work to improve gene annotation is ongoing [1]. Identification of disease causing genes in dogs can provide valuable information about a similar disease process in humans. Dog colonies harboring a specific disease can be developed for study and gene therapy and experimental treatments for a disease are much more amenable in dogs than in humans. This reiterates the benefits of the canine model organism for not only disease mapping, but for functional research of identified genetic variants that may also have human relevance.

In the following thesis studies, the genetic mapping tools and strategies previously described are used to investigate two different genetic diseases in purebred dogs. The first being the complex disease of atopic dermatitis, a pruritic allergic skin disease, in the West Highland White Terrier breed. The second, a simple autosomal recessive neurodegenerative disease, hereditary ataxia, in the Old English Sheepdog breed.

Chapter Aims

The aim of Chapter 2 was to identify a genetic locus segregating with atopic dermatitis in families of West Highland White Terriers using a genome-wide linkage screen with microsatellite markers.

The aim of Chapter 3 was to identify a genomic region or regions associated with atopic dermatitis in West Highland White Terriers and in Old English Sheepdogs with hereditary ataxia using a genome-wide association study.

The aim of Chapter 4 was to use targeted capture of the genomic region associated with hereditary ataxia in Old English Sheepdogs identified in Chapter 3 and sequence the region using next generation sequencing technology to ultimately map the disease causing gene.

The final chapter, Chapter 5, summarizes the conclusions of Chapters 2, 3, and 4 and presents the limitations and future directions of the research.

References

1. Ostrander EA, Ruvinsky A: **The genetics of the dog.** *Electronic resource* 2012, Oxforshire**:** CAB International.

2. [http://omia.angis.org.au]

3. Ostrander EA, Giniger E,: **Semper fidelis: what man's best friend can teach us about human biology and disease.** Am J Hum Genet 1997, **61**(3):475.

4. Ostrander EA, Kruglyak L: **Unleashing the canine genome.** Genome Res 2000, **10**(9):1271-1274.

5. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ,3rd, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, DeJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin CW, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, Grabherr M, Kellis M, Kleber M, Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli KP, Parker HG, Pollinger JP, Searle SM, Sutter NB, Thomas R, Webber C, Baldwin J, Abebe A, Abouelleil A, Aftuck L, Ait-Zahra M, Aldredge T, Allen N, An P, Anderson S, Antoine C, Arachchi H, Aslam A, Ayotte L, Bachantsang P, Barry A, Bayul T, Benamara M, Berlin A, Bessette D, Blitshteyn B, Bloom T, Blye J, Boguslavskiy L, Bonnet C, Boukhgalter B, Brown A, Cahill P, Calixte N, Camarata J, Cheshatsang Y, Chu J, Citroen M, Collymore A, Cooke P, Dawoe T, Daza R, Decktor K, DeGray S, Dhargay N, Dooley K, Dooley K, Dorje P, Dorjee K, Dorris L, Duffey N, Dupes A, Egbiremolen O, Elong R, Falk J, Farina A, Faro S, Ferguson D, Ferreira P, Fisher S, FitzGerald M, Foley K, Foley C, Franke A, Friedrich D, Gage D, Garber M, Gearin G, Giannoukos G, Goode T, Goyette A, Graham J, Grandbois E, Gyaltsen K, Hafez N, Hagopian D, Hagos B, Hall J, Healy C, Hegarty R, Honan T, Horn A, Houde N, Hughes L, Hunnicutt L, Husby M, Jester B, Jones C, Kamat A, Kanga B, Kells C, Khazanovich D, Kieu AC, Kisner P, Kumar M, Lance K, Landers T, Lara M, Lee W, Leger JP, Lennon N, Leuper L, LeVine S, Liu J, Liu X, Lokyitsang Y, Lokyitsang T, Lui A, Macdonald J, Major J, Marabella R, Maru K, Matthews C, McDonough S, Mehta T, Meldrim J, Melnikov A, Meneus L, Mihalev A, Mihova T, Miller K, Mittelman R, Mlenga V, Mulrain L, Munson G, Navidi A, Naylor J, Nguyen T, Nguyen N, Nguyen C, Nguyen T, Nicol R, Norbu N, Norbu C, Novod N, Nyima T, Olandt P, O'Neill B, O'Neill K, Osman S, Oyono L, Patti C, Perrin D, Phunkhang P, Pierre F, Priest M, Rachupka A, Raghuraman S, Rameau R, Ray V, Raymond C, Rege F, Rise C, Rogers J, Rogov P, Sahalie J, Settipalli S, Sharpe T, Shea T, Sheehan M, Sherpa N, Shi J, Shih D, Sloan J, Smith C, Sparrow T, Stalker J, Stange-Thomann N, Stavropoulos S, Stone C, Stone S, Sykes S, Tchuinga P, Tenzing P, Tesfaye S, Thoulutsang D, Thoulutsang Y, Topham K, Topping I, Tsamla T, Vassiliev H, Venkataraman V, Vo A, Wangchuk T, Wangdi T,

Weiand M, Wilkinson J, Wilson A, Yadav S, Yang S, Yang X, Young G, Yu Q, Zainoun J, Zembek L, Zimmer A, Lander ES: **Genome sequence, comparative analysis and haplotype structure of the domestic dog.** Nature 2005, **438**(7069):803-819.

6. King MC, Lee GM, Spinner NB, Thomson G, Wrensch MR: **Genetic epidemiology.** Annu Rev Public Health 1984, **5**:1-52.

7. Haines JL, Pericak-Vance MA: **Genetic analysis of complex diseases.** *Electronic resource 2006,* Hoboken, N.J.: Wiley-Liss.

8. Nussbaum RL: **Thompson & Thompson genetics in medicine.** Philadelphia: Saunders/Elsevier; 2007.

9. Falconer DS, Mackay TF: **Introduction to quantitative genetics.** Essex, England: Longman; 1996.

10. Zondervan KT, Cardon LR: **Designing candidate gene and genome-wide case-control association studies.** Nat Protoc 2007, **2**(10):2492-2501.

11. Nicholas FW: **Simple segregation analysis: a review of its history and terminology.** J Hered 1982, **73**(6):444-450.

12. Ruvinsky A, Sampson J: **The genetics of the dog.** 2001, Wallingford**:** CABI Publishing.

13. Evans DM, Purcell S: **Power calculations in genetic studies.** Cold Spring Harb Protoc 2012, **2012**(6):664-674.

14. Purcell S, Cherny SS, Sham PC: **Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits.** Bioinformatics 2003, **19**(1):149-150.

15. Schork NJ: **Power calculations for genetic association studies using estimated probability distributions.** Am J Hum Genet 2002, **70**(6):1480-1489.

16. Spielman RS, McGinnis RE, Ewens WJ: **Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM).** Am J Hum Genet 1993, **52**(3):506-516.

17. Karlsson EK, Lindblad-Toh K: **Leader of the pack: gene mapping in dogs and other model organisms.** Nat Rev Genet 2008, **9**(9):713-725.

18. Kropatsch R, Petrasch-Parwez E, Seelow D, Schlichting A, Gerding WM, Akkad DA, Epplen JT, Dekomien G: **Generalized progressive retinal atrophy in the Irish Glen of Imaal Terrier is associated with a deletion in the ADAM9 gene.** Mol Cell Probes 2010, **24**(6):357-363.

19. Wang P, Zangerl B, Werner P, Mauldin EA, Casal ML: **Familial cutaneous lupus erythematosus (CLE) in the German shorthaired pointer maps to CFA18, a canine orthologue to human CLE.** Immunogenetics 2010, .

20. Wilbe M, Jokinen P, Truve K, Seppala EH, Karlsson EK, Biagi T, Hughes A, Bannasch D, Andersson G, Hansson-Hamlin H, Lohi H, Lindblad-Toh K: **Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex.** Nat Genet 2010, **42**(3):250-254.

21. Gulcher J: **Microsatellite markers for linkage and association studies.** Cold Spring Harb Protoc 2012, **2012**(4):425-432.

22. Ostrander EA, Giger U, Lindblad-Toh K: **The dog and its genome.** Cold Spring Harbor, N.Y.**:** Cold Spring Harbor Laboratory Press; 2006.

23. Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K, Wiegand NA, Francisco LV, Gibbs M, Aguirre GD, Ostrander EA: **A linkage map of the canine genome.** Genomics 1997, **46**(3):326-336.

24. Wong AK, Ruhe AL, Dumont BL, Robertson KR, Guerrero G, Shull SM, Ziegle JS, Millon LV, Broman KW, Payseur BA, Neff MW: **A comprehensive linkage map of the dog genome.** Genetics 2010, **184**(2):595-605.

25. Sargan DR, Aguirre-Hernandez J, Galibert F, Ostrander EA: **An extended microsatellite set for linkage mapping in the domestic dog.** J Hered 2007, **98**(3):221-231.

26. Breen M, Hitte C, Lorentzen TD, Thomas R, Cadieu E, Sabacan L, Scott A, Evanno G, Parker HG, Kirkness EF, Hudson R, Guyon R, Mahairas GG, Gelfenbeyn B, Fraser CM, Andre C, Galibert F, Ostrander EA: **An integrated 4249 marker FISH/RH map of the canine genome.** BMC Genomics 2004, **5**:65.

27. Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM, Venter JC: **The dog genome: survey sequencing and comparative analysis.** Science 2003, **301**(5641):1898-1903.

28. Hitte C, Madeoy J, Kirkness EF, Priat C, Lorentzen TD, Senger F, Thomas D, Derrien T, Ramirez C, Scott C, Evanno G, Pullar B, Cadieu E, Oza V, Lourgant K,

Jaffe DB, Tacher S, Dreano S, Berkova N, Andre C, Deloukas P, Fraser C, Lindblad-Toh K, Ostrander EA, Galibert F: **Facilitating genome navigation: survey sequencing and dense radiation-hybrid gene mapping.** Nat Rev Genet 2005, **6**(8):643-648.

29. Thomas R, Duke SE, Bloom SK, Breen TE, Young AC, Feiste E, Seiser EL, Tsai PC, Langford CF, Ellis P, Karlsson EK, Lindblad-Toh K, Breen M: **A cytogenetically characterized, genome-anchored 10-Mb BAC set and CGH array for the domestic dog.** J Hered 2007, **98**(5):474-484.

30. Thomas R, Duke SE, Karlsson EK, Evans A, Ellis P, Lindblad-Toh K, Langford CF, Breen M: **A genome assembly-integrated dog 1 Mb BAC microarray: a cytogenetic resource for canine cancer studies and comparative genomic analysis.** Cytogenet Genome Res 2008, **122**(2):110-121.

31. Boyko AR: **The domestic dog: man's best friend in the genomic era.** Genome Biol 2011, **12**(2):216-2011-12-2-216. Epub 2011 Feb 22.

32. Ott J: *Analysis of human genetic linkage:* Baltimore: Johns Hopkins University Press; 1999.

33. MORTON NE: **Sequential tests for the detection of linkage.** Am J Hum Genet 1955, **7**(3):277-318.

34. Lander E, Kruglyak L: **Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results.** Nat Genet 1995, **11**(3):241-247.

35. Elston RC, Stewart J: **A general model for the genetic analysis of pedigree data.** Hum Hered 1971, **21**(6):523-542.

36. Lander ES, Green P: **Construction of multilocus genetic linkage maps in humans.** Proc Natl Acad Sci U S A 1987, **84**(8):2363-2367.

37. Thompson EA: **MCMC Estimation of Multi-locus Genome Sharing and Multipoint Gene Location Scores.** International Statistical Review 2000, **68**(1):53-73.

38. Yuzbasiyan-Gurkan V, Blanton SH, Cao Y, Ferguson P, Li J, Venta PJ, Brewer GJ: **Linkage of a microsatellite marker to the canine copper toxicosis locus in Bedlington terriers.** Am J Vet Res 1997, **58**(1):23.

39. Acland GM, Ray K, Mellersh CS, Gu W, Langston AA, Rine J, Ostrander EA, Aguirre GD: **Linkage analysis and comparative mapping of canine progressive**

rod-cone degeneration (prcd) establishes potential locus homology with retinitis pigmentosa (RP17) in humans. Proc Natl Acad Sci U S A 1998, **95**(6):3048-3053.

40. Acland GM, Ray K, Mellersh CS, Langston AA, Rine J, Ostrander EA, Aguirre GD: **A novel retinal degeneration locus identified by linkage and comparative mapping of canine early retinal degeneration.** Genomics 1999, **59**(2):134-142.

41. Jonasdottir TJ, Mellersh CS, Moe L, Heggebo R, Gamlem H, Ostrander EA, Lingaas F: **Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs.** Proc Natl Acad Sci U S A 2000, **97**(8):4132-4137.

42. Lingaas F, Aarskaug T, Sletten M, Bjerkas I, Grimholt U, Moe L, Juneja RK, Wilton AN, Galibert F, Holmes NG, Dolf G: **Genetic markers linked to neuronal ceroid lipofuscinosis in English setter dogs.** Anim Genet 1998, **29**(5):371-376.

43. Lin L, Faraco J, Li R, Kadotani H, Rogers W, Lin X, Qiu X, de Jong PJ, Nishino S, Mignot E: **The Sleep Disorder Canine Narcolepsy Is Caused by a Mutation in the Hypocretin (Orexin) Receptor 2 Gene [abstract].** 1999; **98:**365-376.

44. Gordon D, Corwin MB, Mellersh CS, Ostrander EA, Ott J: **Establishing appropriate genome-wide significance levels for canine linkage analyses.** J Hered 2003, **94**(1):1-7.

45. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: **Parametric and nonparametric linkage analysis: a unified multipoint approach.** Am J Hum Genet 1996, **58**(6):1347-1363.

46. Elston RC, Spence AM: **Advances in statistical human genetics over the last 25 years.** Stat Med 2006, **25**(18):3049-3080.

47. Teare DM, Barrett JH: **Genetic linkage studies.** Lancet 2005, **366**(9490):1036-1044.

48. Weeks DE, Lange K: **A multilocus extension of the affected-pedigree-member method of linkage analysis.** Am J Hum Genet 1992, **50**(4):859-868.

49. Abecasis GR, Cherny SS, Cookson WO, Cardon LR: **Merlin--rapid analysis of dense genetic maps using sparse gene flow trees.** Nat Genet 2002, **30**(1):97-101.

50. Soontornvipart K, Wangdee C, Kalpravidh M, Brahmasa A, Sarikaputi M, Temwichitr J, Lavrijsen IC, Theyse LF, Leegwater PA, Hazewinkel HA: **Incidence**

and genetic aspects of patellar luxation in Pomeranian dogs in Thailand. Vet J 2013, **196**(1):122-125.

51. Muller C, Distl O: **Scanning 17 candidate genes for association with primary cataracts in the wire-haired Dachshund.** Vet J 2009, **182**(2):342-345.

52. Slatkin M: **Linkage disequilibrium--understanding the evolutionary past and mapping the medical future.** Nat Rev Genet 2008, **9**(6):477-485.

53. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, Ostrander EA: **Extensive and breed-specific linkage disequilibrium in Canis familiaris.** Genome Res 2004, **14**(12):2388-2396.

54. Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, Ostrander EA, Wayne RK: **Linkage disequilibrium and demographic history of wild and domestic canids.** Genetics 2009, **181**(4):1493-1505.

55. Karlsson E, Baranowska I, Wade C, Salmon Hillbertz NHC, Zody M, Anderson N, Biagi T, Patterson N, Pielberg G, Kulbokas E, Comstock K, Keller E, Mesirov J, von Euler H, Kmpe O, Hedhammar A, Lander E, Andersson G, Andersson L, Lindblad-Toh K: **Efficient mapping of mendelian traits in dogs through genome-wide association.** Nat Genet 2007, **39**(11):1321.

56. Kim RN, Kim DS, Choi SH, Yoon BH, Kang A, Nam SH, Kim DW, Kim JJ, Ha JH, Toyoda A, Fujiyama A, Kim A, Kim MY, Park KH, Lee KS, Park HS: **Genome analysis of the domestic dog (Korean Jindo) by massively parallel sequencing.** DNA Res 2012, **19**(3):275-287.

57. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT: **Basic statistical analysis in genetic case-control studies.** Nat Protoc 2011, **6**(2):121-133.

58. Bush WS, Moore JH: **Chapter 11: genome-wide association studies.** PLoS Comput Biol 2012, **8**(12):e1002822.

59. Lewis CM: **Genetic association studies: design, analysis and interpretation.** Brief Bioinform 2002, **3**(2):146-153.

60. Sasieni PD: **From genotypes to genes: doubling the sample size.** Biometrics 1997, **53**(4):1253-1261.

61. Balding DJ: **A tutorial on statistical methods for population association studies.** Nat Rev Genet 2006, **7**(10):781-791.

62. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E: **Variance component model to account for sample structure in genome-wide association studies.** Nat Genet 2010, **42**(4):348-354.

63. Voight BF, Pritchard JK: **Confounding from cryptic relatedness in case-control association studies.** PLoS Genet 2005, **1**(3):e32.

64. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: **Association mapping in structured populations.** Am J Hum Genet 2000, **67**(1):170-181.

65. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.** Nat Genet 2006, **38**(8):904-909.

66. Patterson N, Price AL, Reich D: **Population structure and eigenanalysis.** PLoS Genet 2006, **2**(12):e190.

67. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES: **A unified mixed-model method for association mapping that accounts for multiple levels of relatedness.** Nat Genet 2006, **38**(2):203-208.

68. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES: **TASSEL: software for association mapping of complex traits in diverse samples.** Bioinformatics 2007, **23**(19):2633-2635.

69. Devlin B, Roeder K, Wasserman L: **Genomic control, a new approach to genetic-based association studies.** Theor Popul Biol 2001, **60**(3):155-166.

70. Devlin B, Roeder K: **Genomic control for association studies.** Biometrics 1999, **55**(4):997-1004.

71. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: **Association mapping in structured populations.** Am J Hum Genet 2000, **67**(1):170-181.

72. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E: **Efficient control of population structure in model organism association mapping.** Genetics 2008, **178**(3):1709-1723.

73. Calboli FC, Sampson J, Fretwell N, Balding DJ: **Population structure and inbreeding from pedigree analysis of purebred dogs.** Genetics 2008, **179**(1):593-601.

74. Laird NM, Lange C: **Family-based designs in the age of large-scale gene-association studies.** Nat Rev Genet 2006, **7**(5):385-394.

75. Laird NM, Horvath S, Xu X: **Implementing a unified approach to family-based tests of association.** Genet Epidemiol 2000, **19 Suppl 1**:S36-42.

76. Excoffier L, Slatkin M: **Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population.** Mol Biol Evol 1995, **12**(5):921-927.

77. Clark AG: **Inference of haplotypes from PCR-amplified samples of diploid populations.** Mol Biol Evol 1990, **7**(2):111-122.

78. Stephens M, Smith NJ, Donnelly P: **A new statistical method for haplotype reconstruction from population data.** Am J Hum Genet 2001, **68**(4):978-989.

79. Niu T, Qin ZS, Xu X, Liu JS: **Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms.** Am J Hum Genet 2002, **70**(1):157-169.

80. Barrett JC: **Haploview: Visualization and analysis of SNP genotype data.** Cold Spring Harb Protoc 2009, **2009**(10):pdb.ip71.

81. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** Bioinformatics 2005, **21**(2):263-265.

82. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practial and Powerful Approach to Multiple Testing.** Journal of the Royal Statistical Society 1995, **57**(1):289.

83. Browning BL: **PRESTO: rapid calculation of order statistic distributions and multiple-testing adjusted P-values via permutation for one and two-stage genetic association studies.** BMC Bioinformatics 2008, **9**:309-2105-9-309.

84. Ioannidis JP, Thomas G, Daly MJ: **Validating, augmenting and refining genome-wide association signals.** Nat Rev Genet 2009, **10**(5):318-329.

Salzmann et al. BMC Genetics 2011, 12:37
http://www.biomedcentral.com/1471-2156/12/37

BMC Genetics

RESEARCH ARTICLE Open Access

Genome-wide linkage study of atopic dermatitis in West Highland White Terriers

Cary A Salzmann[1], Thierry JM Olivry[1,3], Dahlia M Nielsen[2], Judith S Paps[1], Tonya L Harris[1] and Natasha J Olby[1,3*]

Abstract

Background: Canine atopic dermatitis (AD) is a common, heritable, chronic allergic skin condition prevalent in the West Highland White Terrier (WHWT). In canine AD, environmental allergens trigger an inflammatory response causing visible skin lesions and chronic pruritus that can lead to secondary bacterial and yeast infections. The disorder shares many of the clinical and histopathological characteristics of human AD and represents an animal model of this disorder that could be used to further elucidate genetic causes of human AD. Microsatellite markers genotyped in families of WHWTs affected with AD were used to perform a genome-wide linkage study in order to isolate chromosomal regions associated with the disorder.

Results: Blood samples and health questionnaires were collected from 108 WHWTs spanning three families. A linkage simulation using these 108 dogs showed high power to detect a highly penetrant mutation. Ninety WHWTs were genotyped using markers from the Minimal Screening Set 2 (MSS-2). Two hundred and fifty six markers were informative and were used for linkage analysis. Using a LOD score of 2.7 as a significance threshold, no chromosomal regions were identified with significant linkage to AD. LOD scores greater than 1.0 were located in a 56 cM region of chromosome 7.

Conclusions: The study was unable to detect any chromosomal regions significantly linked to canine AD. This could be a result of factors such as environmental modification of phenotype, incorrect assignment of phenotype, a mutation of low penetrance, or incomplete genome coverage. A genome-wide SNP association study in a larger cohort of WHWTs may prove more successful by providing higher density coverage and higher statistical power.

Background

Canine atopic dermatitis (AD) is a common, heritable, chronic allergic skin condition that can cause lifelong morbidity in dogs [1,2]. AD is the ninth most common disorder diagnosed in veterinary practice [3], and certain dog breeds such as the West Highland White Terrier (WHWT), with a prevalence as high as 18.7% [4], are more susceptible to the condition than others [5]. In canine AD, environmental allergens such as house dust mites [6] trigger an inflammatory response leading to the development of erythematous macules and papular lesions [7]. Clinical signs usually manifest between six months and three years of age and the chronic pruritus associated with AD leads to excessive licking, alopecia, hyperpigmentation, scaling, lichenification and secondary bacterial and yeast infections [8]. Diagnosis is based on a set of general criteria including chronic or chronic-relapsing pruritus and dermatitis, classical appearance and distribution of lesions, and family history or breed predisposition and elimination of other potential causes for dermatitis such as parasite infection, flea allergy dermatitis, or adverse food reactions [9]. There are currently no laboratory tests that specifically diagnose canine AD [2]. But secondary criteria used to confirm the diagnosis of AD include the demonstration of allergen specific IgE or positive intradermal test to environmental allergens, as well as response to appropriate clinical therapy with glucocorticoids [9]. Treatment includes allergen specific immunotherapy, glucocorticoids and calcineurin inhibitors such as cyclosporine and tacrolimus [2].

Canine AD shares many of the clinical and histopathological characteristics of human AD and is a good animal model for the disease [10] that could prove useful in uncovering additional causative genes in humans. The gene for

* Correspondence: natasha_olby@ncsu.edu
[1]College of Veterinary Medicine, North Carolina State University, Raleigh, NC 27606, USA
Full list of author information is available at the end of the article

filaggrin (*FLG*), located on human chromosome 1 (HSA 1), has had the most consistent association with AD [11] with two common loss of function mutations initially identified [12], and additional mutations identified subsequently [13]. Filaggrin interacts with intermediate filaments in the cytoskeleton and forms the stratum corneum of the epidermis [14]. The stratum corneum acts as a barrier and locks in essential fluids such as water [15]. However, although *FLG* mutations are associated with human AD, they do not fully account for the disease [12,13]. Investigation of the orthologous canine gene for *FLG*, located on canine chromosome 17 (CFA 17), in a cohort of WHWTs failed to show linkage to canine AD [16] raising the possibility that investigation of the cause of AD in dogs may uncover new candidate genes for the human condition. The study reported here aimed to identify an enriched genetic locus linked to AD in our WHWT pedigrees by testing for linkage to genetic markers across the entire genome. The study was unable to detect any chromosomal regions significantly linked to canine AD in WHWTs using a genome-wide family-based linkage approach.

Methods

Recruitment and Sample Collection

WHWTs affected with AD were recruited through referral from veterinary dermatologists and breed clubs across the United States. Pedigrees were collected and the dogs organized into families. Health questionnaires and blood samples were collected from dogs in families containing a minimum of three generations. All protocols were performed with approval from North Carolina State University's Institutional Animal Care and Use Committee. The health questionnaires enquired about each dog's age, pedigree, allergies (including specific questions about the diagnosis and treatment of skin lesions), as well as any family history of allergic disease (if applicable). One of the study's investigators (TJMO), a board certified veterinary dermatologist, determined each dog's affected status by reviewing their health questionnaire and any additional collected information. The diagnosis of AD was made according to standard methods [2] with fulfillment of all the following criteria: family history of pruritic skin lesions; juvenile onset of a pruritic erythematous skin eruption with a characteristic distribution; exclusion of resembling pruritic skin diseases such as non-atopic food allergies, scabies, and skin infections according to standard diagnostic and therapeutic methods; documentation of hypersensitivity against house dust mites, mold or pollen allergens by intradermal testing or allergen-specific IgE serology; response of skin lesions to anti-inflammatory drugs with evidence of efficacy for treatment of canine AD (e.g. oral or topical glucocorticoids, oral cyclosporine or topical tacrolimus). Dogs were considered normal if they reached four years of age with no evidence of clinical signs of AD. Dogs that could not definitively be classified as "affected" or "normal" based on collected information were classified as of "undetermined" status.

Blood samples were collected into acid citrate dextrose (ACD) tubes and shipped overnight on ice to the laboratory where DNA was extracted using the QIAamp DNA Blood Mini Kit (Qiagen, Valencia, CA). DNA concentrations were measured using a ND-1000 NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE) and stored at -80°C before being genotyped.

Power Study

Prior to genotyping, a power study was performed to determine the ability of our pedigrees to identify a linkage region containing a recessive locus when considering various penetrance models. The assumption of a recessive trait was based on previous work that suggested a dominant trait was unlikely [17]. Two recessive models were considered. The first involved a highly penetrant mutation in which an individual had a 98% chance of exhibiting the disease trait if they were homozygous for the risk allele. There was also a 1% chance of exhibiting the disease due to other causes (other loci or environmental factors). The second model assumed a locus with a smaller effect size; homozygous individuals had a 20% chance of manifesting the trait while other genotypes had a 5% chance of doing so (again reflecting the possibility of other loci or environmental factors). In both cases the frequency of the risk allele at the causal locus was simulated to be 0.2. Once a genetic model was determined, genotypes for the causal locus and three linked markers were generated using the program "markerdrop" from the Morgan (v.2.8.1) software package [18]. The true pedigree structure was used as a template when generating the simulated data. Once the simulated data was generated, it was analyzed using the "lm_bayes" program of the Morgan (v.2.8.1) software package [18]. This program implements a likelihood-based Markov chain Monte Carlo (MCMC) approach designed for linkage analyses of large pedigrees containing inbreeding loops, which the program allows to remain intact [19]. The same analysis technique was used for the real data. As a model-based approach, this analysis tool requires specification of the genetic model, including allele frequencies of the causal locus. For the analysis of the simulated data, as for the real data, the frequency of the risk allele was set to 0.39, chosen to match the frequency determined in other recessive canine diseases [20]. To gauge the sensitivity of the analysis method to incorrect specification of risk allele frequency, we simulated data using a risk allele frequency of 0.2, and performed the analysis assuming a frequency of 0.39 (the risk allele frequency in the analysis of the real data was also assumed to be 0.39, though other values were also tested). The simulation of data and the subsequent analysis step were repeated 100 times, and the resulting LOD scores recorded. Power was estimated as the proportion of runs in which the LOD score achieved or exceeded a 2.7 significance threshold within the simulated linkage region.

Genotyping

Whole genome linkage analysis was performed using 288 microsatellite primer sets (Invitrogen, Carlsbad, CA, USA) from the Minimal Screening Set 2 (MSS-2) [21]. In order to perform multiplex PCR, primers were labelled with one of four fluorescent tags (FAM, VIC, NED and PET). Reaction conditions for each marker were optimized for ease of fragment resolution and to prevent wasted resources. The DNA fragments were visualized using an ABI 3730xl DNA Analyzer (Applied Biosystems, Carlsbad, CA, USA) and analyzed using the GeneMapper v.3.7 software (Applied Biosystems, Carlsbad, CA, USA). Genotypes not consistent with familial relationships were removed from further analysis.

Linkage Analysis

Multipoint linkage analysis was performed using the "lm_bayes" program of the Morgan (v.2.8.1) statistical software package [18]. The trait was modelled as a discrete trait with dogs of undetermined or unknown phenotype assigned a value of unknown.

Chromosomal positions of the heterozygous markers were determined using the online canine genetic linkage map at the University of California, Davis Veterinary Genomics Laboratory (http://www.vgl.ucdavis.edu/dogmap/) and homozygous markers were not included in the linkage analysis. The trait frequency was set to default at 0.61 for the common allele and 0.39 for the risk allele and also altered to assess the effect of different allele frequencies. Marker allele frequencies were calculated using the Case Control tests for Samples with Related Individuals (CC-QLS) [22] and Calculation of Kinship and Inbreeding Coefficients Based on Pedigree Information (KinInbcoef) [23] programs from Mary Sara McPeek's website (http://www.stat.uchicago.edu/~mcpeek/software/index.html). The disease trait was set to default at 98% penetrance, but was lowered to 88% and raised to 99% with no significant changes in the results. LOD scores were calculated for each family separately and then summed across families. Appropriate genome-wide LOD score significance threshold using linkage analysis with canine pedigrees was calculated previously by Gordon et al. [24]. Their results indicated that a threshold of 2.7 yielded a global type-I error of 5% and thus LOD scores greater than 2.7 were considered significant in this study. When chromosomal regions with positive LOD scores were identified, haplotypes of microsatellites in these regions were compared in affected and normal dogs for segregation of homozygosity with affected phenotype. The UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) was also used to search for candidate genes previously associated with AD in these regions.

Results

Sample collection and pedigree analysis

Health questionnaires and pedigrees were collected on 180 related WHWTs. Blood samples were obtained from 108 of these dogs (30 affected, 50 normal, and 28 of undetermined status) including 41 males and 67 females. Using the pedigree information from these WHWTs, three family groups were formed with a minimum of three generations between groups. The smallest of the three families contained nine dogs and the largest family contained 121 dogs. The genotyped WHWTs (22 affected, 40 normal, and 28 of undetermined status) ranged in age from 1 to 16 years, and the average age of affected dogs was 6 years (range 1 year to 13 years) and of normal dogs was 9 years (range 4 years to 16 years).

Power Study

In the simulations for which a highly penetrant gene was generated, a LOD score equal to or greater than 2.7 was detected in 100% of the simulation runs, with a maximum score of 7.5 (Figure 1A). For a gene with low penetrance, none (0%) of the simulation runs resulted in a LOD score equal or greater than 2.7, the maximum being 2.6 (Figure 1B).

Genotyping and Linkage Analysis

Seventy-five WHWTs, spanning the three families were genotyped using the MSS-2 microsatellite makers [21]. Additional dogs were genotyped when preliminary analysis gave positive LOD scores in a region, resulting in 90 of the 108 WHWTs being genotyped on some of the markers. Of the 288 microsatellite markers genotyped, 32 were homozygous and discarded and the remaining 256 were used for linkage analysis. These 256 markers covered the genome with an average inter-marker distance of 8.59 cM, with the largest inter-marker distance being 37.6 cM on CFA 7.

The highest LOD score via linkage analysis was 1.25 on CFA 7, while all other chromosomes had scores less than 0.5. CFA7's peak LOD scores were located between markers FH3972 and FH2973 and spanned approximately 56 cM. Repeated analysis with alteration of both penetrance and risk allele frequency did not result in improved LOD scores.

None of the haplotypes showed segregation between affected and normal dogs in the 56 cM region on CFA 7, but the region did contain the gene for *S100 calcium binding protein A8* (*S100A8*), one of the units of the calprotectin complex present in epidermal cells [25]. This

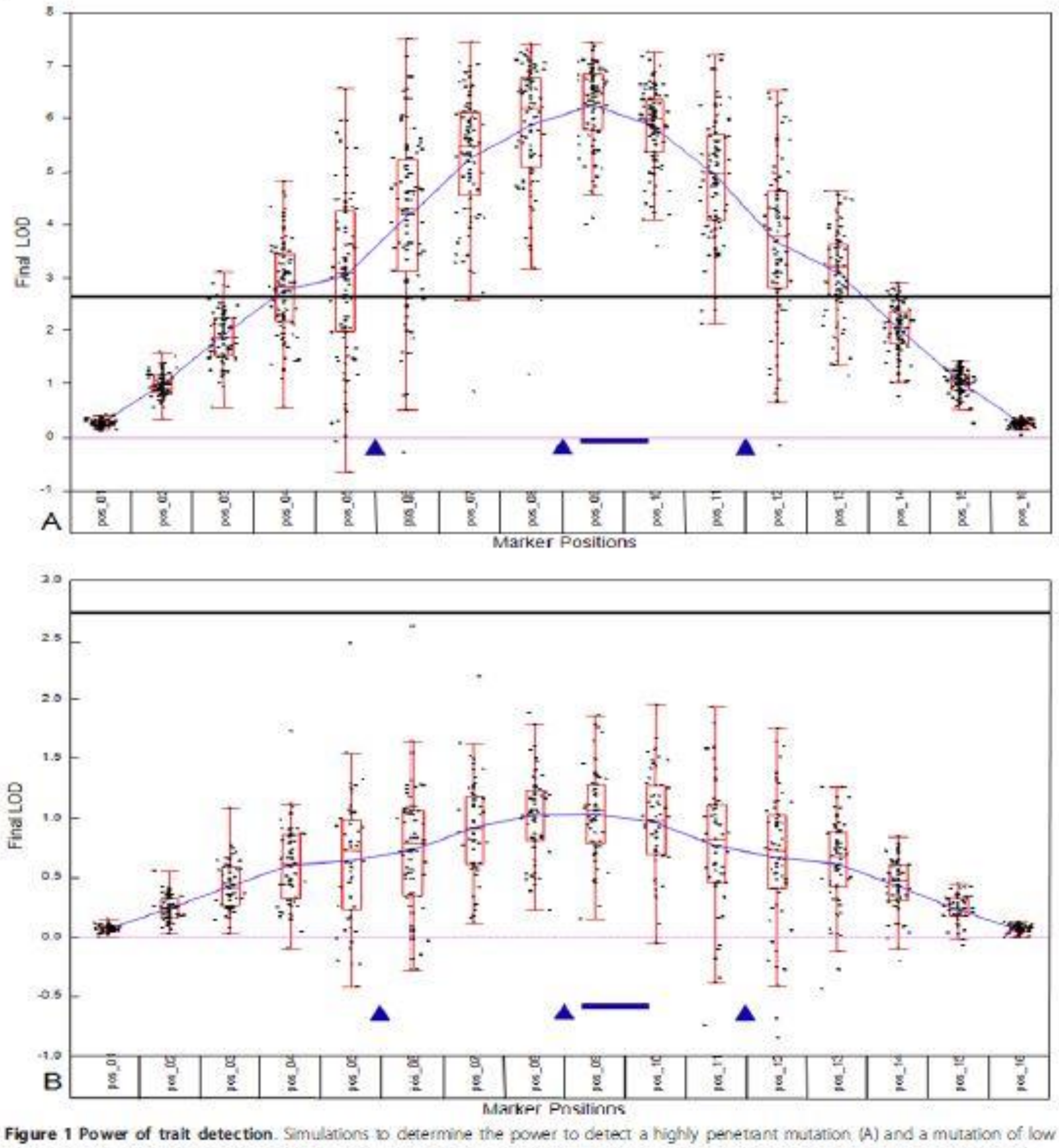


Figure 1 Power of trait detection. Simulations to determine the power to detect a highly penetrant mutation (A) and a mutation of low penetrance (B). The bold horizontal line indicates a LOD score of 2.7, bottom bar between the triangles signifies a mutation segregating with AD and the triangles represent locations of microsatellite markers.

gene has been associated with human and canine AD previously [26-28].

Discussion

This report describes a genome-wide linkage study of canine AD in which no chromosomal regions showed significant evidence for linkage. Genome-wide linkage studies using microsatellites have been a useful tool employed to map disease loci in dogs [29-31] however, their power can be limited in complex traits and in diseases where phenotype is difficult to confirm [32]. Our simulations verified that our linkage study had excellent power to detect a highly penetrant mutation. The failure to identify such a locus could suggest that the inheritance of canine AD in WHWTs is of a more complex nature, as is the case in human AD, which has been described as multifactorial and polygenic with a non-Mendelian mode of inheritance [33]. However, linkage analysis has been used in canine epilepsy in which a recent genome-wide linkage scan in Belgian shepherd dogs identified six loci suggesting association with idiopathic epilepsy, a disease of presumed polygenic mode of inheritance with a recessive gene of major influence [34].

The mode of inheritance of canine AD has been investigated in different breeds but a precise mode of inheritance has not been established [16,17,35,36]. We attempted to determine the mode of inheritance using information on complete WHWT litters from normal parents, yet were extremely limited by the difficulty in obtaining detailed phenotypic information on complete litters. Enumeration of affected and normal dogs in seven complete litters did suggest a major recessive genetic influence. Due to the controlled breeding practices used to generate and perpetuate dog breeds, single gene effects can be enriched within one breed or family, allowing their detection with linkage analysis. However, this proved not to be the case with the WHWT pedigrees we worked with.

Genome-wide association studies (GWAS) using single nucleotide polymorphisms (SNPs) are an increasingly popular approach used to identify genetic mutations involved in complex diseases. A GWAS performed in Golden Retrievers with AD identified two SNPs associated with AD located in intronic regions [37]. A further study by the same group to evaluate SNPs in AD candidate genes was performed in Golden Retrievers and seven other dog breeds including the WHWT [38]. A SNP was identified in the *thymic stromal lymphpoietin receptor* (*TSLPR*) gene that was associated with AD in all eight breeds as well as a SNP in *FLG* associated with AD in UK Labradors. Results of this study suggest that different breeds, as well as dogs of the same breed from different geographic regions have different genetic risk factors for AD [38]. As we did not genotype markers on the X chromosome, we cannot exclude the possibility of linkage to a region including *TSLPR* in our WHWT family.

Although a specific chromosomal region linked to AD in WHWTs was not identified, we were able to exclude linkage to the *FLG* locus, corroborating the results of other studies [16,38].

Linkage could not be ruled out in a 56 cM region on CFA 7. The gene *S100A8*, located within this 56 cM region on CFA 7, has been reported to play a role in canine AD [27,28]. Members of the S100 family are pro-inflammatory molecules released by phagocytes [39] and also keratinocytes (epidermal cells) [40] and are present in the human epidermal differentiation complex (EDC) which has been both linked to and shown association with AD in humans [41]. The three microsatellite markers genotyped in the region were separated by 37.6 and 18.7 cM. Genotyping of additional markers in this 56 cM region may definitively rule-out linkage in the region.

Determination of phenotype may also have influenced the results of our study. Affected dogs were diagnosed by one veterinary dermatologist (TJMO) using stringent criteria thereby enhancing the chance of homozygous diagnoses. However, the assignment of normal phenotype is more likely to represent a source of error; it is possible that dogs classified as normal had minimal or seasonal symptoms that were missed by their owners and veterinarians. It is also possible that an apparently normal dog may not have experienced (and possibly never will) the relevant allergenic triggers. While this is a source of concern, normal phenotype was assigned only to dogs older than four years of age and when possible, owners were re-contacted to confirm that there was no change in affection status over the course of the study which spanned several years. Since AD is usually diagnosed before the age of three, this approach limited erroneous identification of normal dogs.

Finally, although the MSS-2 set of microsatellites provides markers across the entire canine genome, homozygosity of some markers in our population of dogs meant that some regions were not covered at high density. A GWAS in a larger cohort of WHWTs may prove more successful by providing higher density coverage of the entire genome and higher statistical power.

Conclusions

A genetic locus linked to AD in WHWTs was not identified. This could be a result of factors such as environmental modification of phenotype, incorrect assignment of phenotype, a mutation of low penetrance, or incomplete genome coverage.

Acknowledgements
We would like to thank the Westie Foundation of America and the American Kennel Club Canine Health Foundation for financial support as well as the WHWT owners and breeders who made this study possible.

Author details
[1]College of Veterinary Medicine, North Carolina State University, Raleigh, NC 27606, USA. [2]Department of Genetics, North Carolina State University, Raleigh, NC 27606, USA. [3]Center for Comparative Medicine and Translational Research, North Carolina State University, Raleigh, NC 27606, USA.

Authors' contributions
TJMO and NJO conceived the study and NJO directed the study. JSP contacted owners and breeders, coordinated sample collection, organized and managed pedigrees and paperwork and TJMO assigned phenotypes. TLH and CAS performed laboratory work and linkage analysis. DMN performed the power studies and provided analysis support. CAS and NJO wrote the manuscript with contributions from DMN. All authors read and approved the final manuscript.

Received: 31 October 2010 Accepted: 21 April 2011
Published: 21 April 2011

References

1. Olivry T, DeBoer DJ, Griffin CE, Halliwell RE, Hill PB, Hillier A, Marsella R, Sousa CA: **The ACVD task force on canine atopic dermatitis: forewords and lexicon.** *Vet Immunol Immunopathol* 2001, 81(3-4):143-146.
2. Olivry T, DeBoer DJ, Favrot C, Jackson HA, Mueller RS, Nuttall T, Prelaud P, International Task Force on Canine Atopic Dermatitis: **Treatment of canine atopic dermatitis: 2010 clinical practice guidelines from the International Task Force on Canine Atopic Dermatitis.** *Vet Dermatol* 2010, 21(3):233-248.
3. Lund EM, Armstrong PJ, Kirk CA, Kolar LM, Klausner JS: **Health status and population characteristics of dogs and cats examined at private veterinary practices in the United States.** *J Am Vet Med Assoc* 1999, 214(9):1336-1341.
4. Grayson JK: **West Highland White Terrier Health Survey: Final Report.** Westie Foundation of America, Inc; 2007 [http://www.westiefoundation.org/surveys/wfa-health-survey-2006-2007.pdf].
5. Sousa CA, Marsella R: **The ACVD task force on canine atopic dermatitis (II): genetic factors.** *Vet Immunol Imnopathmuol* 2001, 81(3-4):153-157.
6. Hill PB, DeBoer DJ: **The ACVD task force on canine atopic dermatitis (IV): environmental allergens.** *Vet Immunol Immunopathol* 2001, 81(3-4):169-186.
7. Marsella R, Olivry T, Nicklin C, Lopez J: **Pilot investigation of a model for canine atopic dermatitis: environmental house dust mite challenge of high-IgE-producing beagles, mite hypersensitive dogs with atopic dermatitis and normal dogs.** *Vet Dermatol* 2006, 17(1):24-35.
8. Griffin CE, DeBoer DJ: **The ACVD task force on canine atopic dermatitis (XIV): clinical manifestations of canine atopic dermatitis.** *Vet Immunol Immunopathol* 2001, 81(3-4):255-269.
9. DeBoer DJ, Hillier A: **The ACVD task force on canine atopic dermatitis (XV): fundamental concepts in clinical diagnosis.** *Vet Immunol Immunopathol* 2001, 81(3-4):271-276.
10. Girolomoni G, Marsella R: **Canine Models of Atopic Dermatitis: A Useful Tool with Untapped Potential.** *J Invest Dermatol* 2009, 129(10):2351-2357.
11. O'Regan G, Sandilands A, McLean WHI, Irvine A: **Filaggrin in atopic dermatitis.** *J Allergy Clin Immunol* 2009, 124(3 Suppl 2):R2.
12. Palmer CNA, Irvine A, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee S, Goudie D, Sandilands A, Campbell L, Smith FJD, O'Regan G, Watson R, Cecil J, Bale S, Compton J, DiGiovanna J, Fleckman P, Lewis-Jones S, Arseculeratne G, Sergeant A, Munro C, El Houate B, McElreavey K, Halkjaer L, Bisgaard H, Mukhopadhyay S, McLean WHI: **Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis.** *Nat Genet* 2006, 38(4):441-446.
13. Henderson J, Northstone K, Lee SP, Liao H, Zhao Y, Pembrey M, Mukhopadhyay S, Smith GD, Palmer CN, McLean WH, Irvine AD: **The burden of disease associated with filaggrin mutations: a population-based, longitudinal birth cohort study.** *J Allergy Clin Immunol* 2008, 121(4):872-7.e9.
14. Steinert PM, Cantieri JS, Teller DC, Lonsdale-Eccles JD, Dale BA: **Characterization of a class of cationic proteins that specifically interact with intermediate filaments.** *Proc Natl Acad Sci USA* 1981, 78(7):4097-4101.
15. Candi E, Schmidt R, Melino G: **The cornified envelope: a model of cell death in the skin.** *Nature Reviews. Molecular Cell Biology* 2005, 6(4):328-340.
16. Roque JB, O'Leary CA, Kyaw-Tanner M, Latter M, Mason K, Shipstone M, Vogelnest L, Duffy DL: **Haplotype sharing excludes canine orthologous Filaggrin locus in atopy in West Highland White Terriers.** *Anim Genet* 2009, 40(5):793-794.
17. DeBoer HM: **Serum immunoglobulin E concentrations in West Highland White Terrier puppies do not predict development of atopic dermatitis.** *Vet Dermatol* 1999, 10(4):275-281.
18. Thompson EA: **MCMC Estimation of Multi-locus Genome Sharing and Multipoint Gene Location Scores.** *International Statistical Review* 2000, 68(1):53-73.
19. Thompson E, Wijsman E, George A: **MCMC multilocus lod scores: application of a new approach.** *Hum Hered* 2005, 59(2):98-108.
20. Olby N, Blot S, Thibaud JL, Phillips J, O'Brien DP, Burr J, Berg J, Brown T, Breen M: **Cerebellar cortical degeneration in adult American Staffordshire Terriers.** *J Vet Intern Med* 2004, 18(2):201-208.
21. Murphy K, Galibert F, Ostrander E, Guerra T, Williams D, Steiner J, Tsai K, Clark L: **Chromosome-specific microsatellite multiplex sets for linkage studies in the domestic dog.** *Genomics* 2004, 84(3):550-554.
22. Bourgain C, Hoffjan S, Nicolae R, Newman D, Steiner L, Walker K, Reynolds R, Ober C, McPeek M: **Novel case-control test in a founder population identifies P-selectin as an atopy-susceptibility locus.** *Am J Hum Genet* 2003, 73(3):612-626.
23. Karigl G: **A recursive algorithm for the calculation of identity coefficients.** *Ann Hum Genet* 1981, 45(3):299-305.
24. Gordon D, Corwin MB, Mellersh CS, Ostrander EA, Ott J: **Establishing appropriate genome-wide significance levels for canine linkage analyses.** *J Hered* 2003, 94(1):1-7.
25. Ehrchen JM, Sunderkotter C, Foell D, Vogl T, Roth J: **The endogenous Toll-like receptor 4 agonist S100A8/S100A9 (calprotectin) as innate amplifier of infection, autoimmunity, and cancer.** *J Leukoc Biol* 2009, 86(3):557-566.
26. Barnes K: **An update on the genetics of atopic dermatitis: scratching the surface in 2009.** *J Allergy Clin Immunol* 2010, 125(1):16-29.
27. Merryman-Simpson AE, Wood SH, Fretwell N, Jones PG, McLaren WM, McEwan NA, Clements DN, Carter SD, Ollier WE, Nuttall T: **Gene (mRNA) expression in canine atopic dermatitis: microarray analysis.** *Vet Dermatol* 2008, 19(2):59-66.
28. Chung TH, Oh JS, Lee YS, Kang KS, Jung JW, Youn HY, Hwang CY: **Elevated serum levels of S100 calcium binding protein A8 (S100A8) reflect disease severity in canine atopic dermatitis.** *J Vet Med Sci* 2010, 72(6):693-700.
29. Yuzbasiyan-Gurkan V, Blanton SH, Cao Y, Ferguson P, Li J, Venta PJ, Brewer GJ: **Linkage of a microsatellite marker to the canine copper toxicosis locus in Bedlington terriers.** *Am J Vet Res* 1997, 58(1):23-27.
30. Lingaas F, Aarskaug T, Sletten M, Bjerkas I, Grimholt U, Moe L, Juneja RK, Wilton AN, Galibert F, Holmes NG, Dolf G: **Genetic markers linked to neuronal ceroid lipofuscinosis in English setter dogs.** *Anim Genet* 1998, 29(5):371-376.
31. Jonasdottir TJ, Mellersh CS, Moe L, Heggebo R, Gamlem H, Ostrander EA, Lingaas F: **Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs.** *Proc Natl Acad Sci USA* 2000, 97(8):4132-4137.
32. Dukes-McEwan J, Jackson I: **The promises and problems of linkage analysis by using the current canine genome map.** *Mammalian genome* 2002, 13(11):667-672.
33. Cao Y, Liao M, Huang X, Mo Z, Gao F: **Meta-analysis of genome-wide linkage studies of atopic dermatitis.** *Dermatitis* 2009, 20(4):193-199.
34. Oberbauer AM, Belanger JM, Grossman DI, Regan KR, Famula TR: **Genome-wide linkage scan for loci associated with epilepsy in Belgian shepherd dogs.** *BMC Genet* 2010, 11:35.
35. Schwartzman RM, Massicot JG, Sogn DD, Cohen SG: **The atopic dog model: report of an attempt to establish a colony.** *Int Arch Allergy Appl Immunol* 1983, 72(2):97-101.
36. Schwartzman RM: **Immunologic studies of progeny of atopic dogs.** *Am J Vet Res* 1984, 45(2):375-378.
37. Wood SH, Ke X, Nuttall T, McEwan N, Ollier WE, Carter SD: **Genome-wide association analysis of canine atopic dermatitis and identification of disease related SNPs.** *Immunogenetics* 2009, 61(11-12):765-772.
38. Wood SH, Ollier WE, Nuttall T, McEwan NA, Carter SD: **Despite identifying some shared gene associations with human atopic dermatitis the use of multiple dog breeds from various locations limits detection of gene associations in canine atopic dermatitis.** *Vet Immunol Immunopathol* 2010, 138(3):193-197.
39. Roth J, Vogl T, Sorg C, Sunderkotter C: **Phagocyte-specific S100 proteins: a novel group of proinflammatory molecules.** *Trends Immunol* 2003, 24(4):155-158.
40. Eckert RL, Broome AM, Ruse M, Robinson N, Ryan D, Lee K: **S100 proteins in the epidermis.** *J Invest Dermatol* 2004, 123(1):23-33.
41. Barnes KC: **Gene-environment and gene-gene interaction studies in the molecular genetic analysis of asthma and atopy.** *Clinical and experimental allergy* 1999, 29(Suppl 4):47.

doi:10.1186/1471-2156-12-37
Cite this article as: Salzmann *et al.*: **Genome-wide linkage study of atopic dermatitis in West Highland White Terriers.** *BMC Genetics* 2011 12:37.

CHAPTER 3

Genome-wide Association Analyses in West Highland White Terriers with Atopic Dermatitis and in Old English Sheepdogs with Hereditary Ataxia

Introduction

A genome-wide association study (GWAS) most often uses a common form of genetic variation, the single nucleotide polymorphism (SNP), which is a single base change in the DNA sequence. While a single base pair can theoretically be one of four bases, a SNP is typically one of two bases [1]. Thousands to hundreds of thousands of SNPs are assessed across the entire genome; the current Illumina CanineHD BeadChip uses 173,662 SNPs. While there are over 2.5 million known canine SNPs, not all of them need to be genotyped to assess the entire genome due to linkage disequilibrium. Linkage disequilibrium (LD) is the non-random association of alleles at different loci in a population. The LD of a genome gives information on population history, the breeding system and geographic subdivision, while regional LD reflects natural selection, mutation, gene conversion, and forces affecting gene frequencies [2]. Reduction in LD is dependent on recombination. The extent of LD is considered in genome-wide association studies in order to capture information on as much of the genome as possible with as few markers as possible. Instead of using SNPs chosen based on LD, the Illumina CanineHD BeadChip uses evenly spaced SNPs, with approximately 70 SNPs per Mb. The BeadChip consists of 50-mer DNA oligos attached to wells of a silicon slide. Whole genome DNA is amplified and fragmented and the DNA fragments are applied to a well on the slide allowing the DNA to hybridize to the oligos. Once the genomic fragment has hybridized, the oligo primer is extended a single-base at the 3'end. Fluorophores are added such that if the extended base is a C or G, it is labeled with biotin; A or T bases are labeled with dinitrophenyl. A laser excites the fluorophore attached to the extended base, and an image of the emitted light is captured and used to determine the particular allele.

Once the numerous SNPs are genotyped, each SNP needs to be tested for association with the trait of interest. In case-control association studies, the SNPs are analyzed using a contingency table or logistic regression and association can be tested for by comparing the frequencies of each allele of the SNP or the genotype classes against the trait to determine whether there is a difference in frequencies between cases and controls [1, 3]. In addition to counting the genotypes, different genotypic models such as recessive, dominant, multiplicative and additive can be considered and tested [1]. The dominant model considers having one or more copies of an allele to increase risk for the trait; whereas, in the recessive model, having two copies of an allele increases risk [1]. In multiplicative and additive genetic models, the risk for the trait increases by a factor in heterozygotes with that factor doubled in the additive model and squared for homozygotes in the multiplicative model [3]. The PLINK toolset [4] was designed to handle the large datasets that are generated in a GWAS and to address the confounding issues that can occur in such large datasets, a few such issues are described below.

Because hundreds of thousands to millions of SNPs are being tested for association with a trait, a correction for multiple testing must be done to account for the increased likelihood of a false positive result with so many individual tests. A Bonferroni correction does this by adjusting the false positive rate by dividing it by the number of tests performed. However, a Bonferroni correction is conservative because it assumes the each test is independent when in fact some SNPs are correlated due to LD. Another method to correct for multiple tests is permutation testing. In permutation testing, the case and control status of the samples are reshuffled and association tests performed, the final SNP p-values are adjusted based on the number of times the permuted data has a test statistic greater than the original dataset [5].

An additional factor that should be considered when performing a GWAS is population structure, which can also lead to false positive associations. Population structure can be caused by population stratification: individuals coming from different populations [6], or cryptic relatedness: individuals being closely related and therefore not independent samples [7]. Genomic control [8, 9] is one way to account for population structure by calculating the test statistics at SNP loci unlikely to have an effect on the trait and determining an inflation factor then using that inflation factor to adjust the test statistics [10]. A more recent method of efficient mixed-model association (EMMAX) accounts for population structure by using an expedited linear mixed model and calculated kinship matrix from the SNP data to account for relatedness between pairs of individuals to modify association results [6, 11]. While genomic control and EMMAX are useful tools to correct for population structure, a family-based association study can eliminate the need for these corrections. The family-based design is robust against population structure because it tests for both linkage and association between the SNP and phenotype [12].

The transmission disequilibrium test (TDT) uses a family-based design and surveys for a difference in allele transmission from a heterozygous parent and their affected offspring [13]. The family-based association tests (FBATs) are an extension of the TDT method with one such extension being implemented in the FBAT toolkit [14]. This toolkit improves upon the TDT method by handling factors such as: missing parents, additional family members, and different genetic models and phenotypes [14].

The aim of the work presented in this chapter was to identify SNPs associated with atopic dermatitis in West Highland White Terriers and with hereditary ataxia in Old English Sheepdogs. The genetics of these diseases are quite different; atopic dermatitis is inherited in a complex manner with an environmental influence as described in Chapter 2 while hereditary ataxia is a simple autosomal recessive neurodegenerative disease. For both breeds, SNPs spanning the entire canine genome were genotyped using Illumina BeadChips. Genome-wide association analyses were performed with a case-control and family-based design using the genotype data from the breeds.

<u>Material and Methods</u>

West Highland White Terriers and atopic dermatitis

Sample Collection

West Highland White Terriers (WHWT) were recruited through advertisement on the breed society webpage (http://www.westieclubamerica.com/), referral from veterinary dermatologists across the U.S., and from clients being seen in the dermatology clinic at North Carolina State University's College of Veterinary Medicine. Additional WHWT DNA samples collected for a previous linkage study (described in Chapter 2; [15]) were used. WHWT affected with atopic dermatitis were diagnosed according to standard criteria [16] and as reported previously in Chapter 2 [15]. The relevant clinical data was obtained in the form of a questionnaire which the owner or veterinarian filled out (**Figures 1-2**) along with the review of records of any additional clinical tests performed. Multi-generational pedigrees were provided by owners when available.

NC STATE UNIVERSITY
COLLEGE OF VETERINARY MEDICINE

Atopic Dermatitis in West Highland White Terriers
Clinical Questionnaire (One per Westie)

1. Veterinarian's Information: *Date:*
Name:
Mailing Address:
Phone Number:
Email Address:

2. Owner's Information:
Name:
Mailing Address:
Phone Number:
Email Address:
Does the owner agree to be contacted regarding the dog's illness or its origin? Yes ☐ No ☐

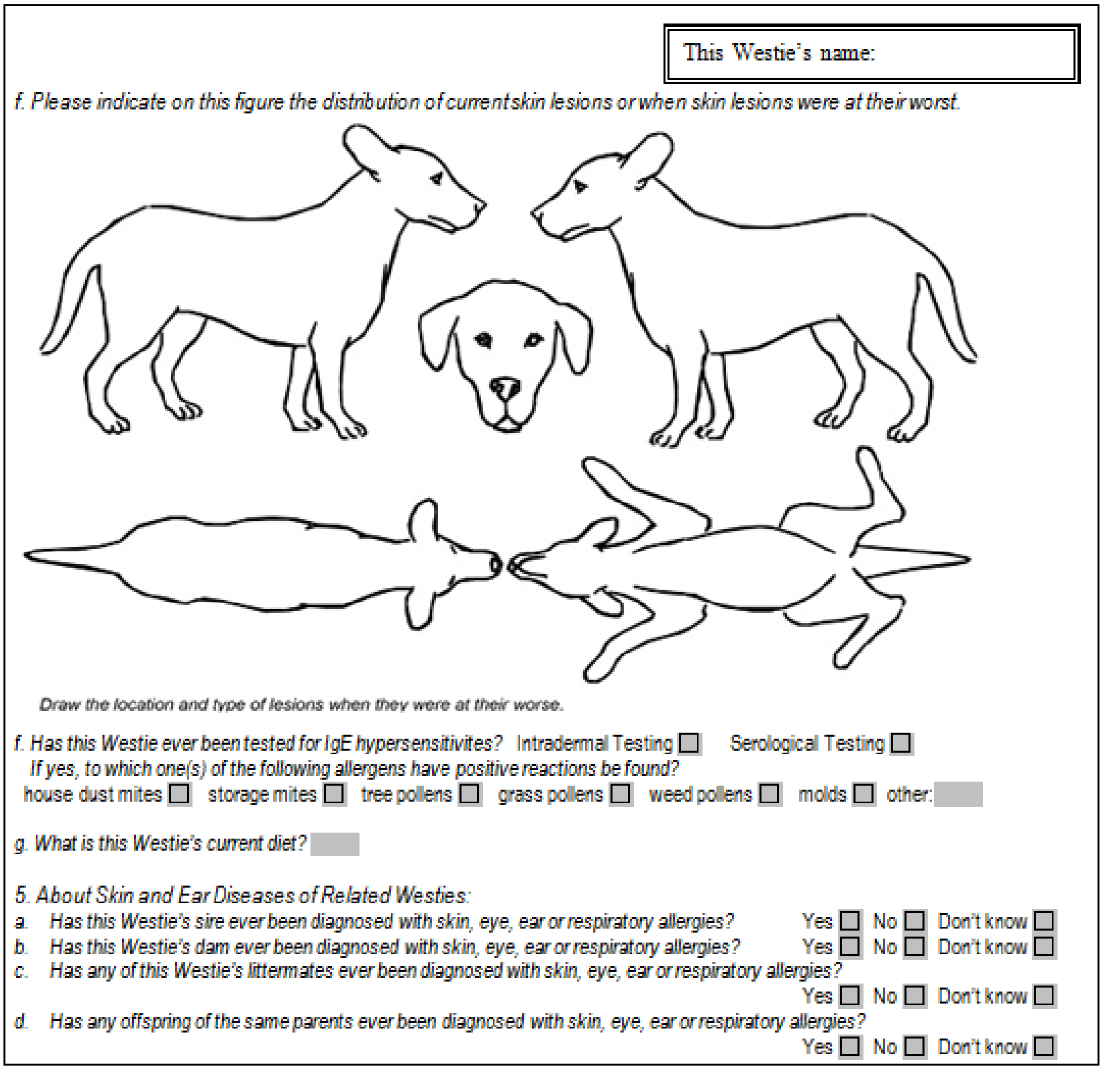
3. Westie's Information:
Westie's Given Name: Male ☐ Female ☐ Neutered/Castrated ☐
Westie's Pedigree Name:
Westie's AKC Registration Number:
Westie's UKC Registration Number (if applicable):
Westie's Date of Birth: Westie's Current Age: years

---

4. Diagnosis of Atopic Dermatitis in this Westie:
a. Is this dog pruritic? Yes ☐ No ☐
b. Have infections been ruled-out based on current standards? Yes ☐ No ☐
b. Have parasitic infestations (e.g. fleas, scabies) been ruled-out based on current standards? Yes ☐ No ☐
c. Have adverse food reactions not manifesting as AD been ruled-out? Yes ☐ No ☐
d. Does this Westie's clinical signs fulfill more than five of the following criteria (Favrot 2009)? Yes ☐ No ☐
(please check all that apply)
- ☐ Onset of clinical signs before three years of age
- ☐ Dog living mostly indoor
- ☐ Glucocorticoid-responsive pruritus (e.g. 0.5 to 1 mg/kg/day of prednisone)
- ☐ Chronic or recurrent Malassezia yeast infections
- ☐ Affected front paws
- ☐ Affected ear pinnae
- ☐ Non-affected ear margins
- ☐ Non-affected dorsolumbar area

e. Is this Westie currently receiving the following drugs (if yes, indicate the daily dose and weight)?
- ☐ oral glucocorticoids (if yes: dosage in mg/kg: ____)
- ☐ injectable glucocorticoids (if yes: dosage in mg/kg: ____)
- ☐ oral cyclosporine (if yes: dosage in mg/kg: ____)
- ☐ cytotoxic drugs (if yes: medication: _____________; dosage in mg/kg: ___)

Figure 1: Page 1 of questionnaire filled out for WHWT with atopic dermatitis.



This Westie's name:

f. Please indicate on this figure the distribution of current skin lesions or when skin lesions were at their worst.

Draw the location and type of lesions when they were at their worse.

f. Has this Westie ever been tested for IgE hypersensitivites? Intradermal Testing ☐ Serological Testing ☐
If yes, to which one(s) of the following allergens have positive reactions be found?
house dust mites ☐ storage mites ☐ tree pollens ☐ grass pollens ☐ weed pollens ☐ molds ☐ other:

g. What is this Westie's current diet?

5. About Skin and Ear Diseases of Related Westies:
a. *Has this Westie's sire ever been diagnosed with skin, eye, ear or respiratory allergies?* Yes ☐ No ☐ Don't know ☐
b. *Has this Westie's dam ever been diagnosed with skin, eye, ear or respiratory allergies?* Yes ☐ No ☐ Don't know ☐
c. *Has any of this Westie's littermates ever been diagnosed with skin, eye, ear or respiratory allergies?* Yes ☐ No ☐ Don't know ☐
d. *Has any offspring of the same parents ever been diagnosed with skin, eye, ear or respiratory allergies?* Yes ☐ No ☐ Don't know ☐

Figure 2: Page 2 of questionnaire filled out for WHWT with atopic dermatitis.

Dogs at least five years old with no signs of atopic dermatitis were considered normal. DNA was extracted from whole blood using QIAamp DNA Blood Midi Kit (Qiagen; Valencia, CA) or from saliva using Oragene^{TM}Animal kit (DNA Genotek; Kanata, Ontario). DNA concentrations were measured using a ND-1000 UV-Vis NanoDrop spectrophotometer (Thermo Scientific; Wilmington, DE). All protocols were performed with approval from North Carolina State University's Institutional Animal Care and Use Committee.

Genotyping and Data Analysis

Ninety-seven (53 male and 44 female) cases and 89 (36 male and 53 female) control dogs were genotyped using Illumina's CanineHD (173,662 SNPs) genotyping BeadChips (Illumina; San Diego, CA). The assays were performed at Gene Seek (Lincoln, NE) according to the manufacturer's instructions and the raw data was returned as IDAT files which were uploaded into Illumina's Genome Studio software (Illumina; San Diego, CA) and analyzed.

Data was pruned using the PLINK toolset v1.07 [4] (http://pngu.mgh.harvard.edu/~purcell/plink/) such that individuals with less than a 95% call rate, SNPs having a minor allele frequency of less than 1%, and missing genotype calls greater than 10% were removed from further analysis. SNP results were checked on the X and Y chromosomes to confirm gender status and as a quality control measure. The pruned dataset was then used to perform case-control association tests using PLINK. The program EMMAX [6] (http://genetics.cs.ucla.edu/emmax/install.html) was also used to help control for any population stratification present. A Bonferroni correction was used to stringently adjust p-values and those less than 0.05 were considered significant. JMP Genomics (SAS; Cary, NC) was used to create a quantile-quantile (Q-Q) plot to determine whether the observed distribution of the test statistics followed the expected distribution and if bias was present in the data. The CanFam2 reference sequence (http://genome.ucsc.edu) was used to investigate genes in the region(s) of the genome in which associations were detected by the GWAS.

Old English Sheepdogs and hereditary ataxia

Sample Collection

Old English Sheepdogs (OES) affected with hereditary ataxia and their family members were recruited through the Old English Sheepdog Club of America, referred from veterinarians, breeders, and owners. Blood or DNA samples and pedigrees were collected. A majority of the samples were collected in the 1990's when the disease first became apparent and problematic in the breed. Affected status was determined by compatible clinical signs of cerebellar disease including ataxia characterized by dysmetria, truncal sway, intention tremors, and spontaneous nystagmus. Additional compatible signs included an age of onset ranging from approximately 6 to 40 months, and a slowly progressive history. When possible, a neurological exam was performed by investigators or a video of the dog was reviewed. Additional tests including cerebrospinal fluid analysis, magnetic resonance imaging of the brain, and post-mortem necropsy and histopathology (considered the gold standard for diagnosis) were undertaken when possible. All protocols were performed with approval from North Carolina State University's Institutional Animal Care and Use Committee. All normal dogs reached four years of age with no evidence of hereditary ataxia. Dogs that could not definitively be classified as "affected" or "normal" based on collected information were classified as an "undetermined" status. DNA was extracted from whole blood using QIAamp DNA Blood Midi Kit (Qiagen; Valencia, CA). DNA concentrations were measured using a ND-1000 UV-Vis NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE).

Genotyping and Data Analysis

Dogs were genotyped on SNPs using Illumina's Canine SNP20 (22,362 SNPs) and CanineHD (173,662 SNPs) genotyping BeadChips (Illumina; San Diego, CA). One control dog was genotyped on both the SNP20 and the CanineHD BeadChips enabling the comparison of genotype calls between them. The assays were performed at the National Institutes of Health's Laboratory of Neurogenetics (Bethesda, MD) according to the manufacturer's instructions. The amplified DNA products were imaged using a BeadArray Reader (Illumina; San Diego, CA) and analyzed using Illumina's Bead Studio and Genome Studio software (Illumina; San Diego, CA).

Data was pruned using PLINK so that individuals with less than a 95% call rate were removed. SNPs having a minor allele frequency of less than 1%, those missing genotype calls greater than 10%, and showing inconsistent calls between the two chips were also removed from further analysis. FBAT [14] (http://www.biostat.harvard.edu/fbat/default.html) was used to check for individuals having a significant number of Mendelian errors. The pruned dataset was then used to perform a case-control analysis (PLINK) and a family-based association analysis (FBAT). A Bonferroni correction was implemented to adjust for multiple testing using PLINK and JMP Genomics (SAS; Cary, NC). P-values less than 0.05 were considered significant. Genotypes were visualized using Microsoft Excel (2010). The UCSC canine genome browser (http://genome.ucsc.edu) was used to identify the genes located in or near to associations revealed by the GWAS.

Results

West Highland White Terriers and atopic dermatitis

DNA samples from a total of 186 (96 cases, 87 controls) WHWT were collected and genotyped on 173,662 SNPs using the Illumina CanineHD BeadChip. The cases' ages ranged from one to fifteen years old (mean=6.11yrs, SD= 3.39 yrs) while the controls' age range was from 5 to 16 years old (mean= 8.70 yrs, SD=2.44 yrs).Twenty-seven of the cases and 44 of the controls had been collected for the linkage study described in Chapter 2. Three dogs were removed during data pruning; one control dog was removed for having a genotyping rate below 95% and one case and one control were removed due to incorrect gender status. The incorrect gender status was believed to have occurred during sample acquisition and not during genotyping; however, the issue could not be resolved; thus, the samples were omitted from further analysis. The association analysis was performed using 96 (52 male and 44 female) cases and 87 (34 male and 53 female) controls and a total data set consisting of 117,576 SNPs after pruning. A case-control analysis using PLINK (**Figure 3**) did not result in any significantly associated SNPs after using a

Bonferroni correction. The top fifteen associated SNPs were present on canine chromosomes 3, 4, 20, 30, 33, and 39 (the X chromosome) (**Table 1**).

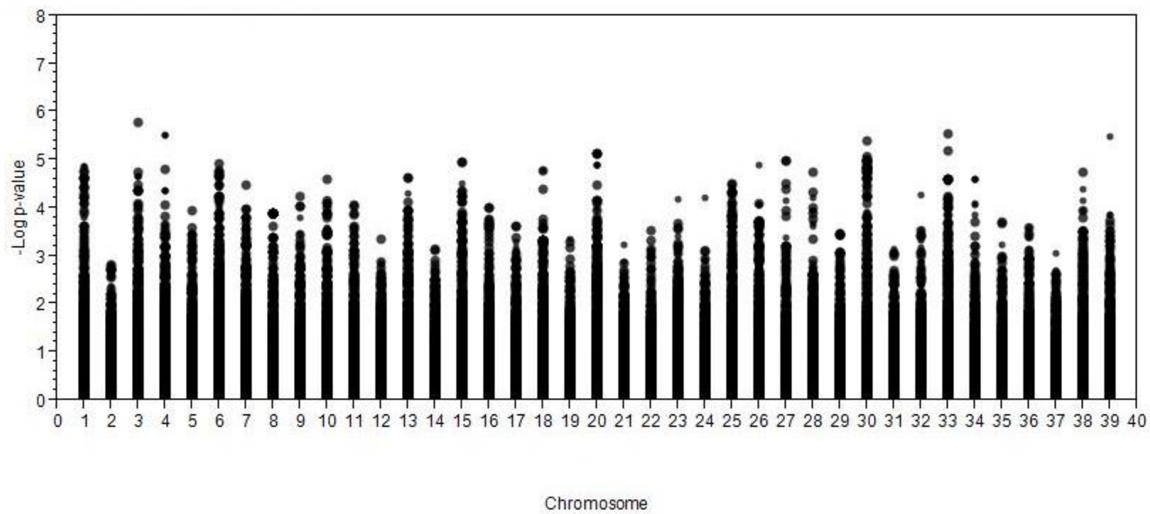


Figure 3: Manhattan skyline plot of unadjusted -log p-values from PLINK against chromosome using 96 WHWT cases, 87 WHWT controls, and 117,576 SNPs.

The genomic inflation factor calculated in PLINK was 1.64, which suggested that population stratification was present in the data and this was verified using a Q-Q plot in JMP. The Q-Q plot of PLINK generated unadjusted p-values shows early separation of the observed from the expected p-values indicating bias in the data most likely caused by population stratification (**Figure 4**).

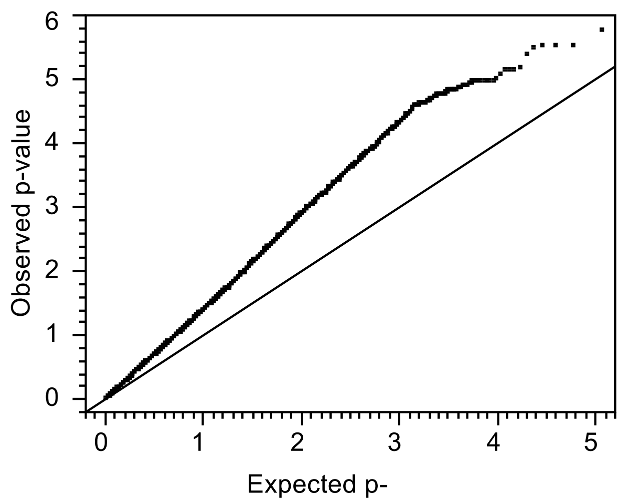


Figure 4: Q-Q plot of unadjusted p-values obtained using PLINK.

In an attempt to control for the population stratification, Emmax was used to perform association tests on the same 117,576 SNPs used in PLINK (**Figure 5**).

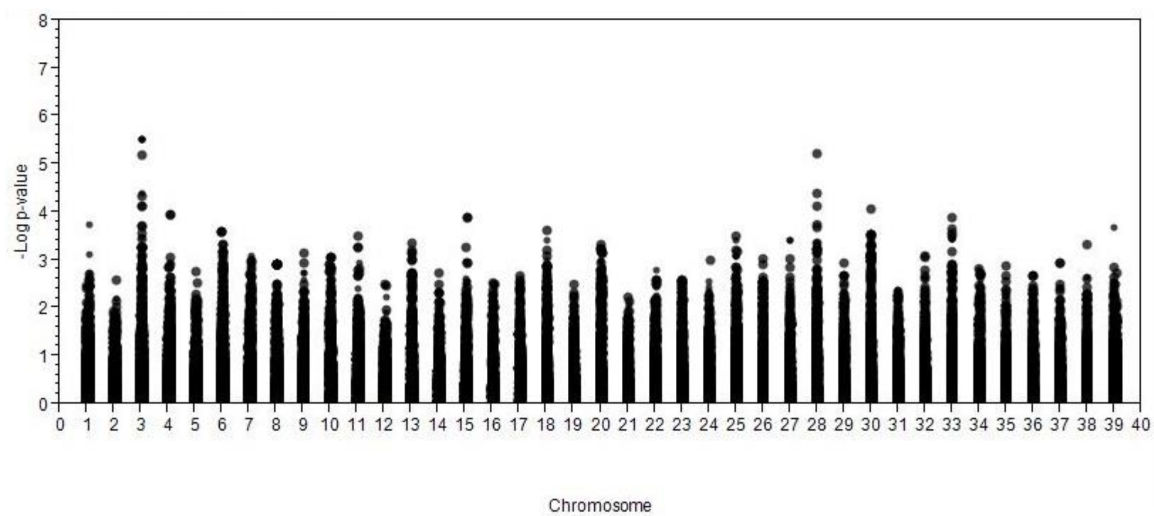


Figure 5: Manhattan skyline plot of unadjusted -log p-values from EMMAX against chromosome using 96 WHWT cases, 87 WHWT controls, and 117,576 SNPs.

Similar to PLINK, none of the associated SNPs passed Bonferroni correction; however, the top 15 associated SNPs from this analysis were present on fewer chromosomes (3, 4, 28, and 30) with over half showing association with a 2.7 Mb region on chromosome 3 (**Table 1**).

Table 1: Top 15 associated SNPs and unadjusted p-values from PLINK and EMMAX. SNPs common to both sets of results are highlighted in pink, yellow, and green.

PLINK			EMMAX		
Chromosome	SNP bp position	P-value	Chromosome	SNP bp position	P-value
3	64,135,951	1.697e-6	3	35,127,520	3.01465e-6
33	6,008,521	2.930e-6	3	35,136,498	3.01465e-6
4	79,231,826	3.021e-6	3	35,146,536	3.01465e-6
4	79,235,632	3.021e-6	3	35,172,960	3.01465e-6
39	21,828,712	3.209e-6	28	15,786,953	6.11351e-6
30	28,410,733	4.134e-6	3	32,423,725	6.52271e-6
33	5,971,017	6.661e-6	3	64,135,951	4.03222e-5
20	41,709,258	7.553e-6	28	15,560,827	4.19362e-5
20	41,745,091	7.553e-6	3	33,710,931	4.68267e-5
20	41,734,129	7.553e-6	3	33,481,012	7.60963e-5
30	28,074,978	8.657e-6	3	33,908,473	7.60963e-5
30	28,131,301	0.00001003	28	16,012,713	0.00007725
30	28,056,658	0.0000105	30	28,410,733	9.00987e-5
30	28,017,721	0.0000105	4	79,231,826	0.00011471
30	28,009,230	0.0000105	4	79,235,632	0.00011471

This 2.7 Mb genomic region contained approximately 37 genes such as *AGGF1*, *CRHBP*, *IQGAP2*, *F2R*, *F2RL1*, *F2RL2*, *MRPL42*, *SAP18*, and *AZNI1* and included 7 olfactory receptor genes. When all nine chromosome 3 SNPs were considered, the region became 31 Mb and determining the number of the genes in the region was difficult. The CanFam2 reference sequence was examined in the two small chromosome 3 regions containing multiple associated SNPs resulting from EMMAX. The SNPs with the smallest p-values were located in a region containing a single gene: *TUBGCP5*; tubulin, gamma complex associated protein 5. The second region

spanning 33,481,012- 33,908,473 bp contained the genes *PSPH*, *POC5*, *EIF3F*, *ANKDD1B*, *POLK*, *HMGCR*, and *COL4A3BP*. The Q-Q plot of the unadjusted p-values generated with results from EMMAX showed fairly tight correlation of observed and expected values indicating the adjustment of population stratification (**Figure 6**).

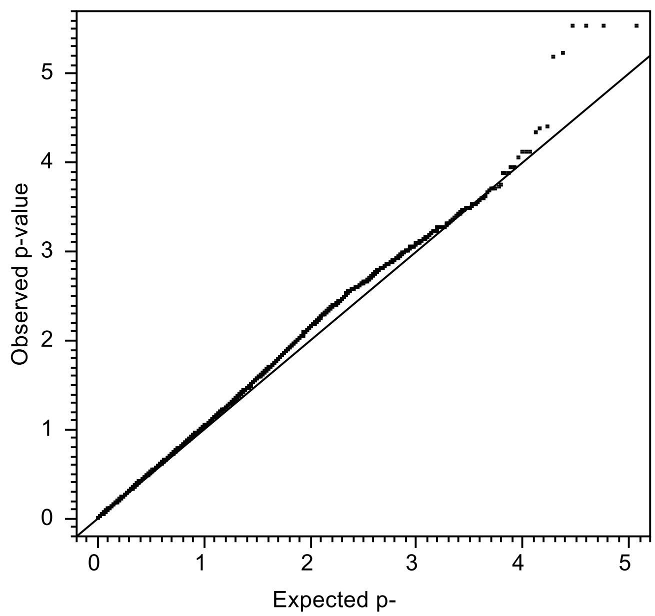


Figure 6: Q-Q plot of unadjusted p-values obtained using EMMAX.

Old English Sheepdogs and hereditary ataxia

DNA samples were collected on 15 cases and 100 controls with most dogs belonging to a large pedigree (**Figure 7**).

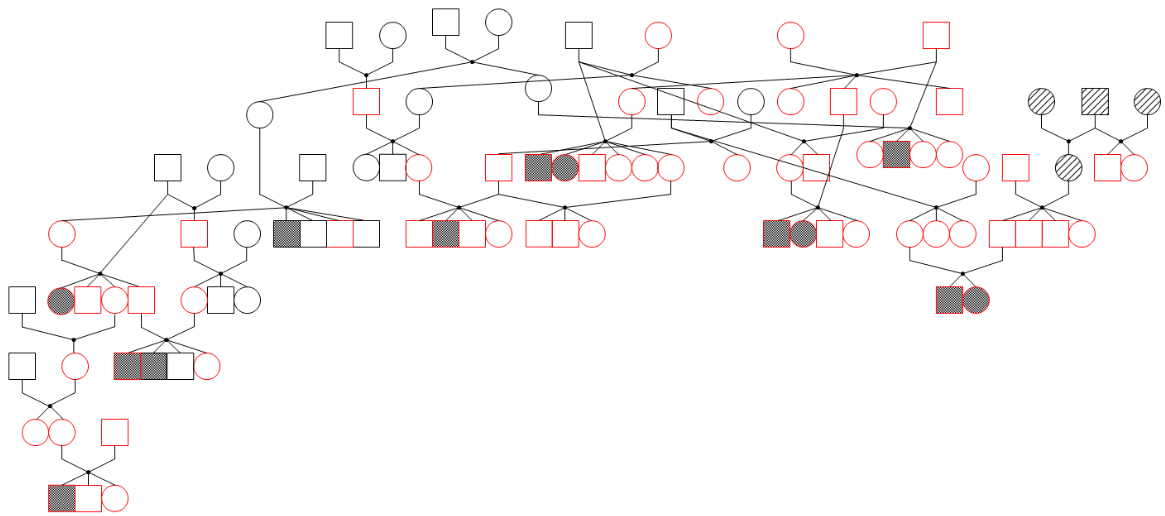


Figure 7: Pedigree of Old English Sheepdogs containing most of the dogs used in the GWAS. Males are represented by squares, females by circles. White shapes are controls, solid shapes are cases, and hashed shapes represent dogs of unknown phenotype. A red outline of the shape indicates DNA was collected on the dog.

Fifty-five (14 case, 41 control) dogs were genotyped on Illumina BeadChips. Nine (6 male, 3 female) cases and 13 (6 male, 7 female) control dogs were genotyped on the SNP20 chip and five (4 male, 1 female) cases and 28 (11 male, 17 female) controls were genotyped on the CanineHD chip. One female control dog was genotyped on both chips to compare calls between the two. All genotyped dogs had a greater than 95% call rate. Genome-wide association analysis was performed using all genotyped dogs which included 14 (10 male, 4 female) case and 40 control (17 male, 23 female) OES.

For the purpose of the GWAS, only SNPs common to both chips were used. After pruning for minor allele frequency, missing genotypes, and inconsistent calls, the final data set consisted of 12,986 SNPs common to both SNP chips. A case-control association analysis using PLINK detected a region of significant association on chromosome 4 with two SNPs passing Bonferroni correction at positions 42,075,319 bp (p_{raw}=1.8x10^{-8}, $p_{Bonferroni}$=1.32x10^{-4}) and 42,030,138 bp (p_{raw}=6.68x10^{-8}, $p_{Bonferroni}$=8.67x10^{-4}) (**Figure 8**).

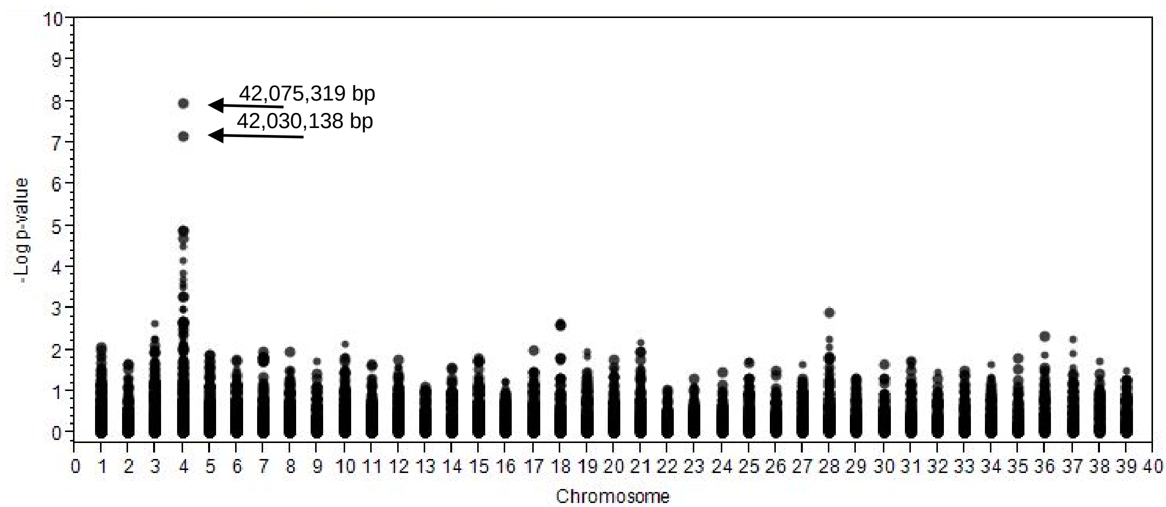


Figure 8: Manhattan skyline plot of unadjusted -log p-values from PLINK against chromosome using 14 cases, 40 controls and 12,986 SNPs. The two SNPs significant after Bonferroni correction are indicated by the arrows.

Four additional highly significant SNPs were detected in the nearby region on chromosome 4 spanning 36,247,158 bp to 40,088,807 bp. As most of the dogs genotyped belonged to the same pedigree, a family-based association test was also performed using FBAT in order to control for false positive associations resulting from population stratification. A significant association using FBAT was detected on chromosome 4 with SNPs spanning 34,159,419 bp-42,030,138 bp with the same two SNPs (42,075,319 bp and 42,030,138 bp) having the highest significance as PLINK with a p-value after Bonferroni correction of p=1.1x10^{-3}. These two SNPs were not located in any genes with the nearest gene to the two being *BOD1,* approximately 27 kb away. The next closest gene, CPEB4, was approximately 195 kb away.

Genotype data was visually inspected in Microsoft Excel (2010) for shared regions of homozygosity in affected dogs. There was an identical region of homozygosity extending from 39,245,536 bp to 41,630,772 bp in all affected dogs (**Figure 9**).

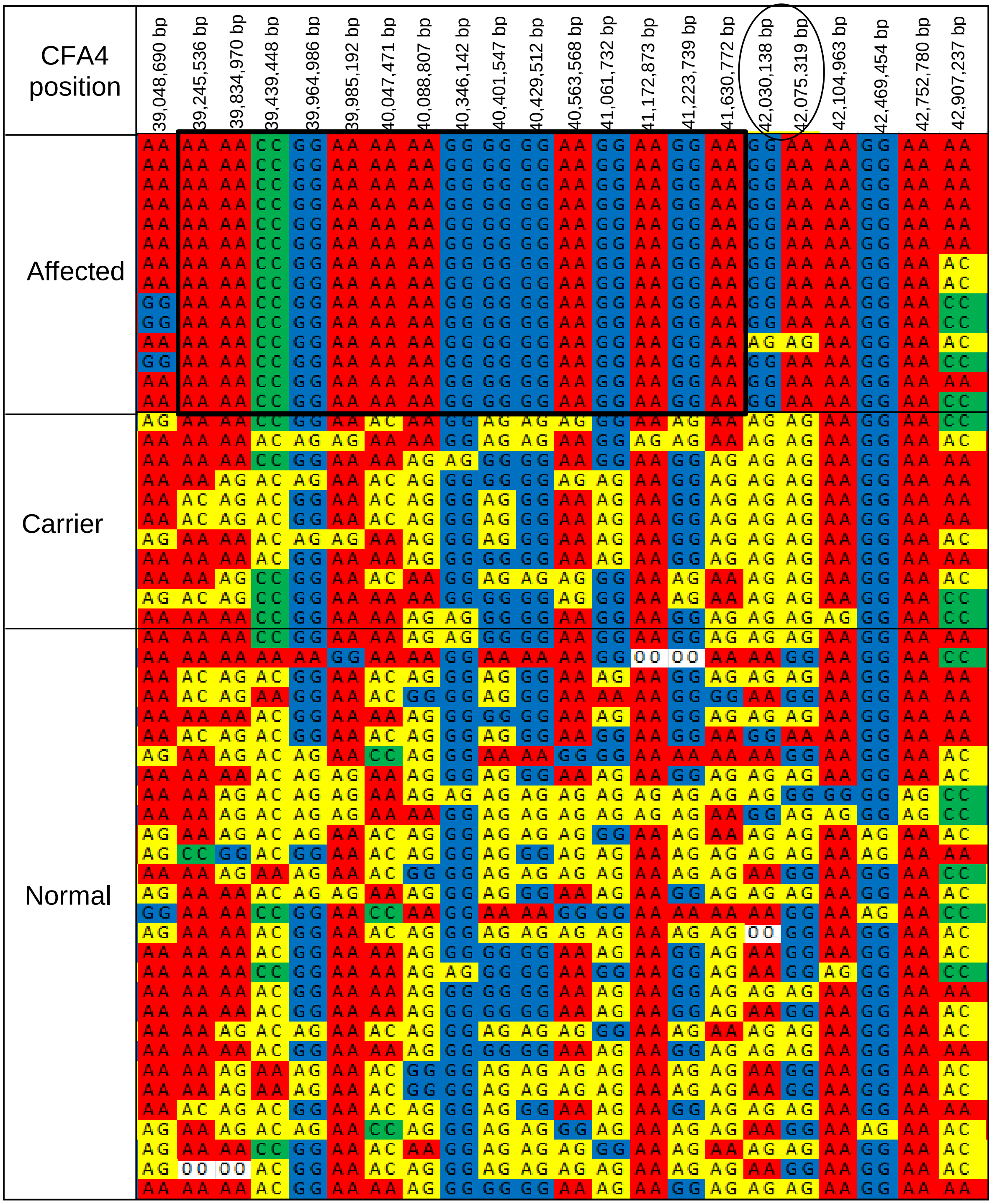


Figure 9: Genotypes of SNPs of Old English Sheepdogs genotyped on Illumina arrays in a region of chromosome 4 (CFA4). Heterozygous genotypes are shaded yellow, homozygous genotypes are blue, red, and green depending on the nucleotide. The white boxes with zeros indicate a missing genotype. The black box

highlights the identical region of homozygosity. The two circled SNPs were significant after Bonferroni correction in analyses using PLINK and EMMAX.

This region was imbedded within a larger relatively homozygous region which extended from approximately 36 Mb to 42 Mb in the affected dogs. Inspection of the canine reference sequence revealed over 30 genes present in the identical 2.4 Mb region and over 80 in the larger 8 Mb region.

Discussion

Genome-wide association analysis using 96 WHWT cases and 87 WHWT controls did not reveal any significantly associated SNPs after correction for multiple testing. The top 15 SNPs resulting from analysis using PLINK and EMMAX did not have a significant overlap of SNPs. While four SNPs: 64,135,951 bp on chromosome 3, 79,235,632 bp and 79,231,826 bp on chromosome 4, and 28,410,733 bp on chromosome 30, were present in the top 15 SNPs from both results, the order in which they were present was not consistent. Similar to the results from PLINK, none of the associated SNPs found using EMMAX passed Bonferroni correction; however, the results from EMMAX were considered more accurate as the Q-Q plot showed non-biased data. There were two regions on chromosome 3 which had multiple SNPs in the top 15 results. The first region spanned 35,127,520 bp-35,172,960 bp, and the second region spanned 33,481,012 bp-33,908,473 bp. The gene *TUBGCP5*; tubulin, gamma complex associated protein 5, is located in the first region spanning 35,127,520 bp-35,172,960 bp. This protein is a subunit of the gamma tubulin complex which localizes to the centrosome and associates with microtubules [17]. Copy number variation involving this gene in humans has been associated with Prader-Willi syndrome [18], Alzheimer's disease [19], and schizophrenia [20, 21]; however, none of these diseases are disorders of the skin. The second region spanning 33,481,012 bp-33,908,473 bp contains genes such as *PSPH*, *POC5*, *EIF3F*, *ANKDD1B*, *POLK*, *HMGCR*, and *COL4A3BP*;

HMGCR being a candidate gene for atopic dermatitis. HMGCR, 3-hydroxy-3-methylglutaryl-CoA reductase, is the rate-limiting enzyme for cholesterol [22]. The synthesis of epidermal cholesterol is involved in maintaining the skin's permeability barrier [23] which is disrupted in atopic dermatitis [24]. The extended 2.7 Mb genomic region from 32,423,725 bp-35,172,960 bp contains the candidate gene *F2RL1*, the gene for protease-activated receptor 2 precursor. This PAR-2 receptor is highly expressed in lesional skin of atopic dermatitis and is involved in homeostasis of the skin's permeability barrier [25].

A few genome-wide association studies of atopic dermatitis have been performed in dog breeds including the WHWT breed. Genome-wide association studies into multiple dog breeds with atopic dermatitis including Boxer, German Shepherd, Labrador and Golden Retriever, Shiba Inu, Shih Tzu, Pit Bull, and WHWT identified three SNPs associated with atopic dermatitis in all eight breeds studied (**Table 2**) [26, 27]. In addition, four SNPs associated with atopic dermatitis in 18 WHWT cases and 48 WHWT controls were identified [26, 27] (**Table 3**).

Table 2: SNPs associated with atopic dermatitis in 8 dog breeds. Highlighted SNPs were present on the Illumina CanineHD chip.

Chromosome	SNP position (bp)	Study	SNP name	Location
10	14,087,517	Wood et al. 2009	rs22114085	intergenic
29	34,162,404	Wood et al. 2009	rs23472497	intergenic
X	751,050	Wood et al. 2010	rs24613432	*TSLPR*

Table 3: SNPs associated with atopic dermatitis in WHWT. Highlighted SNPs were present on the Illumina CanineHD chip.

Chromosome	SNP position (bp)	Study	# cases	# controls
20	5,771,454	Wood et al. 2009	18	48
8	27,285,603	Wood et al. 2009	18	48
20	5,779,740	Wood et al. 2009	18	48
5	56,199,928	Wood et al. 2009	18	48
17	54,773,915	Roque et al. 2011	35	25
17	54,579,514	Roque et al. 2011	35	25
17	54,418,048	Roque et al. 2011	35	25
17	54,792,534	Roque et al. 2011	35	25
1	117,997,337	Roque et al. 2011	35	25
17	54,774,107	Roque et al. 2011	35	25
35	18,843,640	Roque et al. 2011	35	25
17	54,258,071	Roque et al. 2011	35	25
22	61,853,768	Roque et al. 2011	35	25
32	31,930,021	Roque et al. 2011	35	25
14	62,612,290	Roque et al. 2011	35	25
29	40,413,343	Roque et al. 2011	35	25
5	12,547,442	Roque et al. 2011	35	25
17	53,432,468	Roque et al. 2011	35	25
4	8,181,626	Roque et al. 2011	35	25

A later study using 35 WHWT cases and 25 WHWT controls identified significantly associated SNPs in a 1.3 Mb locus on chromosome 17 suggesting polymorphisms in the candidate gene PTPN22 may indicate its role in atopic dermatitis (**Table 3**) [28, 29]. None of the SNPs identified as being associated with atopic dermatitis in WHWT or multiple breeds in these described studies coincide with those found in our GWAS using 96 cases and 87 controls (more than twice the number of dogs than was used in the previous studies) nor were they replicated in the described studies. The CanineHD SNP set used in this dissertation research contained 14 of the 22 SNPs reported previously (**Tables 2-3**). The three previous studies used slightly different genotyping methods including: the SNP20 Illumina BeadChip [27], selected SNPs in candidate genes [26], and the Affymetrix Platinum Panel dog array [28]. All studies corrected for multiple testing however only two of the studies account for population stratification [26, 28]. While affected phenotype assignment was relatively similar in all studies with diagnosis by veterinary dermatologists based on compatible history and clinical signs and exclusion of other causes for pruritus, only Roque el al. described criteria used to assign normal phenotype which included dogs greater than 5 years of age from a non-atopic family. In addition, all the WHWT studies used different cohorts of dogs from different geographical regions and Wood et al. reported different associations in dogs of the same breed from different geographic regions [26]. This study used a WHWT population that came from the US; whereas, the other three studies used WHWT populations from the UK and Australia, which could explain the difference in the associations detected. The filaggrin gene, of which mutations have been identified in human atopic dermatitis, is located on chromosome 17 and has not had associations identified in WHWT; however, associations with atopic dermatitis is UK Labrador Retrievers and a small group of Poodles, Shih Tzus and Pugs have been reported [26, 30]. The two SNPs in *FLG* present in Poodles, Shih Tzus, and Pugs were identified by sequencing of *FLG*; therefore, had not been tested previously in other breeds. While mutations in filaggrin have been associated with atopic dermatitis in some breeds of dogs, our results in the WHWT corroborates other research in the breed in that no associated SNPs were identified near the location of the gene. The CanineHD SNP set did not contain a SNP in the filaggrin gene; however, there was a SNP located within 20 Kb of the gene.

As canine atopic dermatitis is believed to be a complex disease and a GWAS in complex diseases looks for population-wide common variants, there is the possibility that the gene responsible for atopic dermatitis in the breed is not common but rare. The likelihood of detecting a rare variant depends on it being in linkage disequilibrium with a genotyped SNP. The Illumina Canine HD BeadChip uses equally spaced SNPs which may not capture important variation if linkage disequilibrium degrades well within two genotyped SNPs. In addition, to detect a mutation with a small effect such as in complex traits, the necessary number of samples must be genotyped. This number has been estimated to be between 100-300 cases and 100-300 controls for genes conferring a 3-5 fold increase in disease risk [31, 32]. While the population history of purebred dogs suggests diseases with a 5-fold risk are likely, association studies of human atopic dermatitis have been underpowered to detect a moderate affect (odds ratio of 1.5) [31, 33]. The amount of time and resources necessary to collect well phenotyped samples of a single breed exhibiting a complex trait make it difficult to perform a study using 100-300 cases and 100-300 controls. While the phenotyping of the dogs was made by careful review of the dogs history and medical records, it should be noted that control dogs were much more difficult to identify (evidenced by the fewer samples collected). Atopic dermatitis is triggered by environmental allergens and some of the normal dogs may truly be affected dogs who have not been exposed to the environmental trigger. We attempted to minimize this by using normal dogs over the age of five who had a greater chance of encountering the environmental trigger for the disease than a young dog. The complex nature of canine atopic dermatitis, involving many different and possibly interacting genetic and environmental factors may account for the difficulty in identifying significantly associated SNPs in the described GWAS.

The investigation into the autosomal recessive disease of hereditary ataxia in Old English Sheepdogs is an example of how the genetic nature of the disease can impact the ability to achieve significant results.

Genome-wide association analysis in OES identified significant SNPs and an approximately 2.4 Mb homozygous region on CFA4 present in all affected dogs. While the two most significant SNPs, which passed a stringent Bonferroni correction, were just outside of the 2.4 Mb homozygous region and additional significant SNPs were present between 34 Mb and 42 Mb, it was difficult to determine where to begin searching for candidate genes. There are over 80 genes located in this 8Mb region of which none were clear candidates for hereditary ataxia. The UCSC genome browser did not indicate that the top two SNPs at 42,075,319 bp and 42,030,138 bp were located in any genes and the two nearest genes were *BOD1* and *CPEB4*.

BOD1, biorientation of chromosomes in cell division 1, is a centrosomal and kinetochore protein necessary for the proper orientation of chromosomes in the cell [34]. Cytoplasmic polyadenylation element binding protein 4 (CPEB4) is one member of a family of proteins responsible for polyadenylation-induced protein translation in early development. A PubMed database search did not reveal mutations in *BOD1* or *CPEB4* are associated with any ataxic disease. The time and cost of the resources needed to Sanger sequence the two described genes and the other 80 genes in the associated region were prohibitive. Targeted sequence capture and sequencing using next generation methods can overcome these obstacles and is described in the following chapter.

References

1. Bush WS, Moore JH: **Chapter 11: genome-wide association studies.** PLoS Comput Biol 2012, **8**(12):e1002822.

2. Slatkin M: **Linkage disequilibrium--understanding the evolutionary past and mapping the medical future.** Nat Rev Genet 2008, **9**(6):477-485.

3. Lewis CM: **Genetic association studies: design, analysis and interpretation.** Brief Bioinform 2002, **3**(2):146-153.

4. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** Am J Hum Genet 2007, **81**(3):559-575.

5. Browning BL: **PRESTO: rapid calculation of order statistic distributions and multiple-testing adjusted P-values via permutation for one and two-stage genetic association studies.** BMC Bioinformatics 2008, **9**:309-2105-9-309.

6. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E: **Variance component model to account for sample structure in genome-wide association studies.** Nat Genet 2010, **42**(4):348-354.

7. Voight BF, Pritchard JK: **Confounding from cryptic relatedness in case-control association studies.** PLoS Genet 2005, **1**(3):e32.

8. Devlin B, Roeder K, Wasserman L: **Genomic control, a new approach to genetic-based association studies.** Theor Popul Biol 2001, **60**(3):155-166.

9. Devlin B, Roeder K: **Genomic control for association studies.** Biometrics 1999, **55**(4):997-1004.

10. Balding DJ: **A tutorial on statistical methods for population association studies.** Nat Rev Genet 2006, **7**(10):781-791.

11. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E: **Efficient control of population structure in model organism association mapping.** Genetics 2008, **178**(3):1709-1723.

12. Laird NM, Lange C: **Family-based designs in the age of large-scale gene-association studies.** Nat Rev Genet 2006, **7**(5):385-394.

13. Spielman RS, McGinnis RE, Ewens WJ: **Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM).** Am J Hum Genet 1993, **52**(3):506-516.

14. Laird NM, Horvath S, Xu X: **Implementing a unified approach to family-based tests of association.** Genet Epidemiol 2000, **19 Suppl 1**:S36-42.

15. Salzmann CA, Olivry TJ, Nielsen DM, Paps JS, Harris TL, Olby NJ: **Genome-wide linkage study of atopic dermatitis in West Highland White Terriers.** BMC Genet 2011, **12**:37-2156-12-37.

16. Olivry T, International Task Force of Canine Atopic Dermatitis: **New diagnostic criteria for canine atopic dermatitis.** Vet Dermatol 2010, **21**(1):123-126.

17. Murphy SM, Preble AM, Patel UK, O'Connell KL, Dias DP, Moritz M, Agard D, Stults JT, Stearns T: **GCP5 and GCP6: two new members of the human gamma-tubulin complex.** Mol Biol Cell 2001, **12**(11):3340-3352.

18. Bittel DC, Kibiryeva N, Butler MG: **Expression of 4 genes between chromosome 15 breakpoints 1 and 2 and behavioral outcomes in Prader-Willi syndrome.** Pediatrics 2006, **118**(4):e1276-83.

19. Ghani M, Pinto D, Lee JH, Grinberg Y, Sato C, Moreno D, Scherer SW, Mayeux R, St George-Hyslop P, Rogaeva E: **Genome-wide survey of large rare copy number variants in Alzheimer's disease among Caribbean hispanics.** G3 (Bethesda) 2012, **2**(1):71-78.

20. Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, Muglia P, Francks C, Matthews PM, Gylfason A, Halldorsson BV, Gudbjartsson D, Thorgeirsson TE, Sigurdsson A, Jonasdottir A, Jonasdottir A, Bjornsson A, Mattiasdottir S, Blondal T, Haraldsson M, Magnusdottir BB, Giegling I, Moller HJ, Hartmann A, Shianna KV, Ge D, Need AC, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Paunio T, Toulopoulou T, Bramon E, Di Forti M, Murray R, Ruggeri M, Vassos E, Tosato S, Walshe M, Li T, Vasilescu C, Muhleisen TW, Wang AG, Ullum H, Djurovic S, Melle I, Olesen J, Kiemeney LA, Franke B, GROUP, Sabatti C, Freimer NB, Gulcher JR, Thorsteinsdottir U, Kong A, Andreassen OA, Ophoff RA, Georgi A, Rietschel M, Werge T, Petursson H, Goldstein DB, Nothen MM, Peltonen L, Collier DA, St Clair D, Stefansson K: **Large recurrent microdeletions associated with schizophrenia.** Nature 2008, **455**(7210):232-236.

21. Zhao Q, Li T, Zhao X, Huang K, Wang T, Li Z, Ji J, Zeng Z, Zhang Z, Li K, Feng G, St Clair D, He L, Shi Y: **Rare CNVs and Tag SNPs at 15q11.2 Are Associated With Schizophrenia in the Han Chinese Population.** Schizophr Bull 2012, .

22. Rodwell VW, Nordstrom JL, Mitschelen JJ: **Regulation of HMG-CoA reductase.** Adv Lipid Res 1976, **14**:1-74.

23. Feingold KR, Man MQ, Menon GK, Cho SS, Brown BE, Elias PM: **Cholesterol synthesis is required for cutaneous barrier function in mice.** J Clin Invest 1990, **86**(5):1738-1745.

24. Marsella R, Sousa CA, Gonzales AJ, Fadok VA: **Current understanding of the pathophysiologic mechanisms of canine atopic dermatitis.** J Am Vet Med Assoc 2012, **241**(2):194-207.

25. Lee SE, Jeong SK, Lee SH: **Protease and protease-activated receptor-2 signaling in the pathogenesis of atopic dermatitis.** Yonsei Med J 2010, **51**(6):808-822.

26. Wood SH, Ollier WE, Nuttall T, McEwan NA, Carter SD: **Despite identifying some shared gene associations with human atopic dermatitis the use of multiple dog breeds from various locations limits detection of gene associations in canine atopic dermatitis.** Vet Immunol Immunopathol 2010, .

27. Wood SH, Ke X, Nuttall T, McEwan N, Ollier WE, Carter SD: **Genome-wide association analysis of canine atopic dermatitis and identification of disease related SNPs.** Immunogenetics 2009, **61**(11-12):765-772.

28. Roque JB, O'Leary CA, Duffy DL, Kyaw-Tanner M, Gharahkhani P, Vogelnest L, Mason K, Shipstone M, Latter M: **Atopic dermatitis in West Highland white terriers is associated with a 1.3-Mb region on CFA 17.** Immunogenetics 2012, **64**(3):209-217.

29. Roque JB, O'Leary CA, Kyaw-Tanner M, Duffy DL, Gharahkhani P, Vogelnest L, Mason K, Shipstone M: **PTPN22 polymorphisms may indicate a role for this gene in atopic dermatitis in West Highland white terriers.** BMC Res Notes 2011, **4**:571-0500-4-571.

30. Suriyaphol G, Suriyaphol P, Sarikaputi M, Theerawatanasirikul S, Sailasuta A: **Association of Filaggrin (*FLG*) Gene Polymorphism with Canine Atopic Dermatitis in Small Breed Dogs.** Thai J Vet Med 2011, **41**(4):509-510-517.

31. Karlsson EK, Lindblad-Toh K: **Leader of the pack: gene mapping in dogs and other model organisms.** Nat Rev Genet 2008, **9**(9):713-725.

32. Karlsson E, Baranowska I, Wade C, Salmon Hillbertz NHC, Zody M, Anderson N, Biagi T, Patterson N, Pielberg G, Kulbokas E, Comstock K, Keller E, Mesirov J, von Euler H, Kmpe O, Hedhammar A, Lander E, Andersson G, Andersson L, Lindblad-Toh K: **Efficient mapping of mendelian traits in dogs through genome-wide association.** Nat Genet 2007, **39**(11):1321.

33. Barnes K: **An update on the genetics of atopic dermatitis: scratching the surface in 2009.** J Allergy Clin Immunol 2010, **125**(1):16.

34. Porter IM, McClelland SE, Khoudoli GA, Hunter CJ, Andersen JS, McAinsh AD, Blow JJ, Swedlow JR: **Bod1, a novel kinetochore protein required for chromosome biorientation.** J Cell Biol 2007, **179**(2):187-197.

CHAPTER 4

Targeted Next Generation Sequencing Reveals a *RAB24* Mutation Segregating with Hereditary Ataxia in Old English Sheepdogs and Gordon Setters

Introduction

While genome-wide association studies have been extremely successful in identifying disease predisposing mutations and predictors of traits in dogs, they do have limitations. Often many genes lie in a region of identified association with no clear candidates presenting themselves. Sequencing the numerous genes would be time consuming as well as expensive using traditional Sanger sequencing [1]. Next generation or second-generation sequencing also called massively parallel or high-throughput sequencing, greatly reduces the time and cost of large scale sequencing by generating billions of short sequence reads with greater coverage of each target base. While multiple technologies have been developed to implement this next generation sequencing, they share some characteristics in common such as: the synthesis of a DNA strand using a polymerase, modified nucleotides, fluorescence detection, and clonal amplification of the template DNA [2]. The Illumina/Solexa technology uses fluorescently labeled reversible terminator nucleotides. As the newly synthesized DNA strand is extended via DNA polymerase, one of the nucleotides is added and synthesis is terminated, un-incorporated nucleotides are removed, and the fluorescence is imaged. The dye and terminator attached to the nucleotide are then removed and a fresh set of nucleotides are added, then the cycle repeats. Prior to sequencing, the template DNA is fragmented, adaptors are ligated to the ends of the fragments which then bind to primers affixed to a solid surface. These fragments are then amplified to form clusters of hundreds of identical DNA fragments which can then be sequenced. The Illumina HiSeq sequencing machine generates 100bp paired end reads with an error rate of approximately 0.1% [3].

When using next-generation sequencing; often, it is not necessary to sequence the entire genome as a region of interest has already been identified. In this case, the region of interest needs to be selected somehow prior to sequencing. A Roche NimbleGen custom-designed microarray enables targeted sequencing by capturing a genomic region(s) using oligo probes attached to a glass slide on which genomic DNA is hybridized and then eluted [4-6]. When comparing target capture methods including solution-based, molecular inversion probes, and microarray; microarray was found to have the greatest percentage of targeted bases associated with high-quality genotypes of the three methods [7].

Once the sequence data has been generated, it needs to be analyzed. The quality of the data is an important factor; thus, a quality score is assigned to each base. The common score is a Phred quality score (Q score), which gives the probability of a base being called incorrectly [8]; a Q score of 30 would indicate an incorrect base call 1 in 1000 times. The sequence reads must next be aligned to a reference genome of which there are numerous software packages available. The Burrows-Wheeler Alignment tool (BWA) is an efficient tool for realignment and supports pair-end mapping of reads generated using Illumina sequencing instruments [9]. Once the sequence reads are aligned, variation must be deciphered which can be done using tools such as The Genome Analysis Toolkit (GATK) [10]. GATK uses the Java programming language and provides users with tools such as variant calling and coverage calculation and has been tested on Illumina generated data [10].

Targeted sequencing of a genomic region identified by GWAS has been successful at identifying disease causing mutations for episodic falling and congenital keratoconjuctivitis sicca and ichthyosiforn dermatosis in the Cavalier King Charles Spaniel [11] as well as a skeletal disease in Brazilian Terriers [12]. Targeted sequencing was also used to sequence a 1.5 Mb genomic region identified by reduced homozygosity in Shar-Pei dogs harboring a mutation for hyaluronanosis [13]. In addition to genomic regions or the entire genome, next generation sequencing technology can be used for sequencing coding exons of the entire genome (exome) and transcripts (RNA-seq). An additional advantage of next generation sequencing has been shown recently in the sequencing of an individual case. An 8bp deletion in the *SPTBN2* gene was identified in a single Beagle diagnosed with neonatal cerebellar cortical degeneration using mRNA-seq [14].

Cerebellar degenerative diseases are the most common type of neurodegenerative diseases in domestic animals [15] and the nervous system is the system most frequently affected by inherited diseases in dogs [16]. Canine hereditary ataxia, also known as cerebellar degeneration, cerebellar abiotrophy, cerebellar cortical degeneration, and cerebellar ataxia results in the spontaneous degeneration of cerebellar tissue [17]. Dogs with the disease display an abnormal gait, called ataxia, which can be described as incoordination. This is evident in their gait where they have difficulty regulating the rate, range and force of movement typically resulting in hypermetria (overstepping). The main pathological consequence of the disease is a loss of Purkinje cells while secondary changes include; reduction of the molecular layer, proliferation of Bergmann astrocytes in areas of Purkinje loss, and swollen Purkinje cell axons [18]. A definitive diagnosis of hereditary ataxia is usually made upon necropsy and histopathology; however, MRI studies have shown that cerebellar atrophy can be detected using morphometric measures: a brainstem to cerebellum mid-sagittal cross-sectional area ratio of 89% or greater is diagnostic [19]. A slowly progressive adult-onset form of the disease consistent with an autosomal recessive mode of inheritance has been described in Old English Sheepdogs [20] with dogs showing typical clinical signs of hypermetric gait, wide-based stance, and ataxia. Hereditary ataxia has also been reported in numerous dog breeds with most breeds having varying disease characteristics such as age of onset, clinical signs, histopathology, and disease progression [21]. A similar age of onset to the Old English Sheepdog occurs in the Gordon Setter [22, 23] and Brittany spaniel breeds [24, 25] along with a mixed breed [26] and Boxer dog [27]. In cases where the inheritance pattern is investigated, autosomal recessive transmission has most commonly been concluded [23, 28-30]. The gene responsible for causing the disease has been identified in three dog breeds thus far: the Coton de Tulear, Finnish Hound, and Beagle. A severe neonatal ataxia consistent with an autosomal recessive mode of inheritance and histological changes that were visible ultrastructurally in the cerebellum was described in the Coton de Tulear breed [30], and a subsequent GWAS identified the causative mutation to be a retrotransposon inserted into exon 8 of *GRM1, a glutamate receptor [31].* The study of an early onset autosomal recessive ataxia with reduced cerebellar size in the Finnish Hound revealed a mutation in *SEL1L*, a component of the ER-associated protein degradation machinery [32]. Sequencing of mRNA followed by a candidate gene search of a young Beagle affected with cerebellar degeneration led to identification of an 8 bp deletion in the *SPTBN2* gene, a gene also associated with human spinocerebellar type 5 [14, 33]. All of the breeds in which a mutation has been identified share an early age of onset which is different from Old English Sheepdogs, making it unlikely that these three genes are responsible for the disease in the breed.

Hereditary ataxias are also common in humans and can be sporadic or inherited. Hereditary ataxias can be autosomal dominant, autosomal recessive, X-linked, and mitochondrial [34]. Some key pathologic themes for the disease include, nucleotide repeat expansions, ion channel dysfunction, defects in DNA repair, and mutations in mitochondrial genes [35]. Approximately 30 autosomal dominant spinocerebellar ataxias (SCAs) have been documented with most having an associated gene or loci identified [36]. Seven SCA subtypes are caused by a CAG (polyglutamine) repeat expansion in specific genes with an additional three types resulting from altered gene expression caused by repeat expansions outside the specific gene’s coding region [37]. Polyglutamine repeats greater than 35-40 assemble into aggregates inside inclusion bodies in the cell [38]. Additional mutated genes responsible for various SCAs include *SPTBN2* [33], *KCNC3* [39], *PRKCG* [40], and *FGF14* [41].

Friedreich's ataxia is the most common autosomal recessive hereditary ataxia to occur in Caucasians [42]. In addition to ataxia; propriopceptive, sensory, and oculomotor disturbances occur as well as vision and hearing loss [43]. This ataxia is caused by an expanded trinucleotide repeat in the first intron of the *FXN* gene on chromosome 9 which reduces *frataxin* expression levels [43]. Ataxia telangiectasia, another autosomal recessive ataxia, is a multisystem disorder with cerebellar ataxia accompanied by immunodeficiency and malignancies along with ocular telangiectasia [44]. Null mutations in the *ATM* gene, a serine-threonine kinase involved in cellular response to DNA damage-specifically double-stranded breaks, are responsible for the disease [45, 46].

Hereditary ataxias have also been described in domestic mammals such as cattle, sheep, horses, cats, and rabbits [17, 47] and numerous naturally occurring: *Lurcher*, *hot-foot*, *staggerer*, *Purkinje cell degeneration*, and *reeler*, and induced ataxic mouse models exist [34]. The Lurcher and hotfoot mouse are examples of ataxia caused by mutations in *GRID2*, a glutamate receptor gene [48, 49]. The heterozygote Lurcher mouse is caused by a gain in malfunction of *GRID2* and has complete loss of Purkinje cells along with secondary granule cell loss [50]. In the hot-foot mouse, mutant alleles 4J and Nancy are caused by deletions in the coding sequence of *GRID2* [49, 51] with Nancy mutants having defective granule-Purkinje innervation and mild granule cell loss [50].The Purkinje cell degeneration (pcd) mouse with defects in the gene *AGTPBP1*/*NNA1* shows complete degeneration of Purkinje cells in adulthood [52]. Nna1 protein was also found to be reduced in hamsters with an ataxic gait [53]. The autosomal recessive staggerer and reeler mice result from mutations in the *RORA* gene [54] and *RELN* respectively [55]. Mouse models exist for many of the autosomal dominant SCAs as well as for both Friedrich's ataxia and Ataxia telangiectasia.

A mutation has been identified in Arabian horses with cerebellar ataxia via linkage analysis and subsequent sequencing of the homozygous identified region [56]. The potentially causative mutation is located in the regulatory region of the

MUTYH (muty Homolog), a DNA glycosylase that removes adenine residues incorrectly added opposite 8-oxo, which is expressed in the brain [56].

In this study, targeted sequence capture and massive parallel sequencing was performed on a 12 Mb region of canine chromosome 4, which is associated with hereditary ataxia in Old English Sheepdogs (OES). Genetic variants were identified and further tested to identify those most likely to be the cause of the disease. Once a candidate mutation had been identified, a larger cohort of OES along with additional dog breeds, including breeds reported to have hereditary ataxia, were screened for the mutation. Quantitative PCR was performed along with sequencing of the transcript to examine changes in gene expression.

Material and Methods

Targeted Sequence Capture

A 12 Mb genomic region of interest was targeted using a custom array designed and manufactured by Roche NimbleGen (Madison, WI). 19,720 tiled probes approximately 60-90bp in length covered approximately 96% of the targeted bases between 34,000,000bp-46,000,000bp on chromosome 4. Unique probes were determined using the Sequence Search and Alignment by Hashing Algorithm (SSAHA) [57]. Targeted bases not captured were due to SSAHA's inability to determine valid probes possibly resulting from non-unique sequence, repetitive sequence, homopolymer runs, or ambiguous bases.

Three micrograms of genomic DNA from three cases and three controls were used to prepare libraries for sequencing. DNA samples were fragmented by sonication (Covaris; Woburn, MA) and the fragments end-repaired, A-tailed, and ligated to indexing oligonucleotide adapters using NEBNext reagents (New England Biolabs; Ipswich, MA). Indexing adapters were provided by the Broad Institute (Cambridge, MA). The indexed DNA fragments were enriched by PCR using AccuPrime (Life Technologies; Grand Island, NY) and Phusion (New England Biolabs, Ipswich, MA) DNA polymerases. Purification was done using QIAquick and

MinElute PCR purification kits (Qiagen; Valencia, CA). Size selection and purification was done using Agencourt AMPure XP beads (Beckman Coulter; Beverly, MA). Analysis and quantification of DNA libraries were done using the Agilent Bioanalyzer DNA 1000 (Agilent Technologies; Santa Clara, CA) and Quant-iT dsDNA HS assay (Life Technologies, Grand Island, NY). Libraries were hybridized onto the array using a NimbleGen hybridization system (Roche NimbleGen; Madison, WI) at 42° C for 70 hours. Arrays were washed and samples eluted using NimbleGen's elution system (Roche NimbleGen; Madison, WI). Post-capture amplification was done using the primers 5'-AATGATACGGCGACCACCGAG-3' and 5'-GAAGCAGAAGACGGCATACGA-3' with Phusion (New England Biolabs; Ipswich, MA) and AccuPrime (Life Technologies; Grand Island, NY) enzymes. Quantitative PCR (qPCR) using QuantiFast® SYBR® Green PCR kit (Qiagen; Valencia, CA) was done to estimate relative fold-enrichment of the target region using primers specific for inside and outside the targeted region of interest (**Table 1**).

Table 1: Forward and reverse primers for quantitative PCR to estimate target region enrichment using a NimbleGen array.

	Forward Primer	Reverse Primer
On target 1	5'-TCATCTCAGCGTTCCACAAG-3'	5'-TGTCATTAGAGGGAGGCAAG-3'
On target 2	5'-CAAGGGGAAGAATGGAAAGA-3'	5'-TAAAGGAGAGTGGGCAGGTT-3'
On target 3	5'-AAGGTGAAGGCAAGCAATGT-3'	5'-TGAGGAGCAGGTGTTGTTCT-3'
Off target 1	5'-TGGCATGTGTCTCCTTTCAA-3'	5'-TCCTGTCCCGTGATTTACAG-3'
Off target 2	5'-ATGGGTCTGGGATGAAGGAT-3'	5'-AGTCTGAGGGCCATTGAACA-3'
Off target 3	5'-TGCAATTTCCCACATGCTC-3'	5'-TTGAGACAGATGGTATGCAGTG-3'

Paired-end sequencing was done on Illumina's Hi-Seq 2000 machine (Illumina; San Diego, CA) at the Broad Institute (Cambridge, MA).

Sequencing Data Analysis

Sequence reads were aligned to the CanFam2.0 canine reference genome using the Burrows-Wheeler Alignment (BWA) tool [9]. SNPs and insertion/deletions (indels) were determined using the Genome Analysis Toolkit (GATK; *http://www.broadinstitute.org/gatk/*) [10]. Sequence data was processed according to the current best practices for data analysis found on the online GATK Wiki page (http://www.broadinstitute.org/gsa/wiki/index.php/Main_Page) and included: initial SNP and indel calling, correcting alignment errors due to indels and inaccurate base quality scores, SNP and indel calling after the realignment and recalibration, and filtering the data using standard filtering parameters. Sequence and variants were viewed with the Integrative Genomic Viewer (IGV) [58] and a Q30 was used as a confidence threshold when considering whether a base was called accurately. Variants were identified in which all three affected dogs were homozygous for the non-reference allele, the two control carriers were heterozygous, and the final control dog was either heterozygous or homozygous for the reference allele. Variants resulting in amino acid changes of coding regions of the genome were considered potential causative mutations for hereditary ataxia.

Re-sequencing and Genotyping Variants

Variants considered as potential causative mutations were genotyped via Sanger sequencing in Old English Sheepdogs and additional dog breeds. The *RAB24* gene was also sequenced in samples collected from dogs affected with hereditary ataxia of various breeds with in which the A>C *RAB24* variant was not present. DNA was extracted from whole blood using QIAamp DNA Blood Midi Kit (Qiagen; Valencia, CA), frozen tissues; DNeasy Blood and Tissue Kit (Qiagen; Valencia, CA), buccal swabs; QIAamp DNA Mini Kit (Qiagen; Valencia, CA), and from saliva using the Oragene^{TM}Animal kit (DNA Genotek; Kanata, Ontario). DNA concentrations were measured using a ND-1000 UV-Vis NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE). Primers were designed using Primer3 (http://frodo.wi.mit.edu/) [59] and NCBI's Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) [60] (**Table 2**).

Table 2: Forward and reverse primers for sequencing of SNPs identified using massively parallel sequencing.

Gene	Forward	Reverse
RGR	5'-CCCTTGTGTCACGACAGGT-3'	5'-CTCCTCTGCGGTGACAGG-3'
RAB24	5'-GTGAGGCAGCCCAGACAG-3'	5'-CCAGGGGACTTACCCAAATA-3'
NSD1	5'-GAGCTGCCTGAAACCAAGAC-3'	5'-TCCCGTTGTAGTGGAACCTC-3'
GPRIN1	5'-TGGCATGTGTCTCCTTTCAA-3'	5'-TCCTGTCCCGTGATTTACAG-3'
CDHR2	5'-AGCTGAGAAATGGTATCCAC-3'	5'-GGGCTTGGTGCTAGAGTCTG-3'

PCR reactions included: 10X Buffer B (Thermo Fisher Scientific; Waltham, MA), 25mM $MgCl_2$ (Thermo Fisher Scientific; Waltham, MA), 10X MasterAmp (Epicentre® Biotechnologies Madison, WI), 25mM dNTPs (Apex BioResearch Products, San Diego, CA), 10µM forward and reverse primer (IDT; Coralville, IA), Taq DNA polymerase (Apex BioResearch Products; San Diego, CA) and molecular grade water to a volume of 27 µL. DNA amounts varied from approximately 25ng to 100ng per reaction. Thermocycler conditions varied by primer set. PCR products were purified using either Agencourt AMPure XP kit (Beckman Coulter; Beverly, MA), QIAquick PCR Purification Kit (Qiagen; Valencia, CA), or MinElute PCR Purification Kit (Qiagen; Valencia, CA). The sequencing of each purified PCR product (20-40 ng/µl for 8µl) was carried out using the same forward and reverse primers used for PCR. Sequencing was performed by Eurofins MWG Operon (Huntsville, AL) following the standard BigDye® Terminator v3.1 manufacture's protocol (Applied Biosystems; Foster, CA) with capillary electrophoresis carried out on the ABI 3730xl DNA Analyzer (Applied Biosystems; Foster, CA).

RNA Sequencing and qRT-PCR

Sections of cerebellum were collected from euthanized affected Old English Sheepdogs (n=2), Gordon Setters (n=2) and normal Beagles (n=2). The tissue samples were immediately frozen in liquid nitrogen after removal from the cerebellum. RNA was extracted from approximately 30 mg of the frozen tissue using Qiagen's RNeasy Mini Kit (Qiagen; Valencia, CA) and concentrations measured using a ND-1000 UV-Vis spectrophotometer (Thermo Scientific; Wilmington, DE).

Reverse transcription was performed using the Stratagene AffinityScript Multiple Temperature cDNA synthesis kit (Agilent Technologies; Santa Clara, CA) and 0.5 µg oligo(dT) and 0.2 µg random primers. Thermocycler conditions were set to 25°C for 10 minutes, 42°C for 5 minutes, 25°C for 60 minutes, and 72°C for 15 minutes.

Primers to amplify the mRNA were designed using Primer3 (http://frodo.wi.mit.edu/) [59] forward primer 5'-CGTGTCTCCAGGCGTAGC-3' and reverse primer 5'-ACTGGGGGTAGCTCAGAC-3'. The approximately 850bp PCR product was excised from an ethidium bromide stained 2% agarose gel and cleaned using the QIAquick Gel Extraction Kit (Qiagen; Valencia, CA). The cDNA of an affected Old English Sheepdog, Gordon Setter, and normal Beagle were sequenced using the same forward and reverse PCR primers. Sequencing was done the same as for genotyping the five SNPs as described in the previous section.

Real-time PCR was performed using Applied Biosystems' StepOne Plus instrument and Invitrogen's Power SYBR Green Master Mix (Life Technologies; Grand Island, NY). *RAB24* primers were designed using Primer3 (http://frodo.wi.mit.edu/) [59] and PerlPrimer (http://perlprimer.sourceforge.net/) [61]. The housekeeping gene *RPS19* (primers published [62]) was used for normalization. 20µl reaction volumes containing 200 nM primers were done in triplicate. Cycling conditions consisted of 95°C for 10 minutes followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute with a melt curve performed at completion. Reaction efficiencies were calculated using a 6 point 2:1 dilution series of 100 ng cDNA mixture. *RAB24* and *RPS19* reaction efficiencies were estimated to be 90% and 88% respectively with both having r^2 values of 0.999. Relative expression differences were calculated using the ΔΔCt-method and statistical significance determined using Microsoft Excel (2010) by performing an unpaired t-test on the relative differences between cases and controls.

Results

Targeted Sequence Capture

Targeted sequence capture was performed to sequence the entire genomic region of homozygosity extending from 34 Mb to 46 Mb on chromosome 4. Six dogs were sequenced, three affected and three normal dogs, which included two obligate heterozygotes and one unrelated, phenotypically normal dog. A Roche NimbleGen array (Madison, WI) was designed to provide coverage of 95.7% of the targeted region using unique probes. The design was reviewed to ensure that all predicted and known gene exons had adequate coverage. Quantitative PCR in genomic regions of prepared libraries inside and outside the target region revealed sufficient capture of the on-target regions to warrant sequencing. Following next generation sequencing on Illumina's Hi-Seq 2000 machine (Illumina; San Diego, CA), approximately 96% of the targeted bases had at least 2X coverage with approximately 72% having at least 30X coverage. The mean coverage of all the samples was determined to be 47X using GATK (**Table 3**).

Table 3: Coverage of the six Old English Sheepdogs sequenced using Illumina's Hi-Seq-2000.

Sample	Total reads	Mean	% Bases >2X	% Bases >10X	% Bases >20X	% Bases >30X
Case 1	623,290,679	51.94	96.1	90.9	83.4	74.2
Case 2	721,627,162	60.14	96.2	91.9	85.9	78.8

Table 3: Continued

Case 3	543,897,169	45.32	95.9	90	81.7	71.3
Control 1	562,824,780	46.9	95.9	90.2	82.3	72.4
Control 2	455,672,762	37.97	95.7	88.7	78.5	65
Control 3	474,255,541	39.52	96	89.4	80.6	69
Average	563,594,682	46.97	96.0	90.2	82.1	71.8

Sequencing Data Analysis

Data processing using the Genome Analysis Toolkit (GATK) revealed 40,711 total variants (32,475 SNPs and 8,236 indels). After filtering the total variants for known SNPs and those present in all six dogs sequenced, 28,061 variants remained, of which only 288 were present in coding regions. Screening of these 288 variants revealed nine SNPs that segregated in an autosomal recessive inheritance pattern; six of which resulted in an amino acid change and were considered potential causative mutations for the disease trait. The six SNPs were located in the genes *RGR, RAB24, NSD1, GPRIN1,* and *CDHR2* (**Table 4**).

Table 4: Details of the six non-synonymous SNPs detected by next generation sequencing of a chromosome 4 (CFA4) region in six Old English Sheepdogs.

CFA4 SNP position (bp)	Gene	Reference allele	Alternate allele	Amino acid change
35,707,658	*RGR*	G	T	Trp>Leu
39,246,812	*RAB24*	A	C	Gln>Pro
39,319,721	*NSD1*	C	G	Lys>Asn
39,929,599	*GPRIN1*	G	A	Arg>Lys
39,929,601	*GPRIN1*	G	A	Ala>Thr
39,940,607	*CDHR2*	C	G	Arg>Ser

Re-sequencing and Genotyping Variants

The six non-synonymous SNPs identified as potential causative mutations were verified by Sanger sequencing in the six dogs initially sequenced. Additional Old English Sheepdogs (a minimum of 2 cases and 11 controls) were genotyped on the six SNPs to determine if they segregated with the phenotype (**Table 5**).

Table 5: Old English Sheepdog genotypes of six non-synonymous SNPs on chromosome 4 (CFA4).

CFA4 SNP position (bp)	Gene	Case phenotype	Case genotype	Control phenotype	Control genotype
35,707,658	*RGR*	5	5-T/T	14	2-T/T, 8-G/T, 4-G/G
39,246,812	*RAB24*	15	15-C/C	100	55-A/C, 45-A/A
39,319,721	*NSD1*	15	15-G/G	100	55-C/G, 45-C/C
39,929,599	*GPRIN1*	9	9-A/A	42	14-A/A, 19-G/A, 9-G/G
39,929,601	*GPRIN1*	9	9-A/A	42	14-A/A, 19-G/A, 9-G/G
39,940,607	*CDHR2*	7	7-G/G	21	4-G/G, 13-C/G, 4-C/C

The SNPs in *RGR, GPRIN1*, and *CDHR2* did not segregate with phenotype in this larger population and were eliminated from further analysis. The two SNPs present in *RAB24* and *NSD1* showed segregation in the cases and controls consistent with an autosomal recessive mode of inheritance. When these two SNPs were tested for association with the trait in the same 54 dogs as the GWAS, the p-values were both highly significant at $p_{Bonferroni}=5.7x10^{-9}$. When additional dog breeds were genotyped on the two SNPs, the *NSD1* SNP was present in the heterozygous state in three neurologically normal Labrador Retrievers and two neurologically normal Standard Poodles; however, only the *RAB24* reference allele was seen in these dogs. Because the *NSD1* SNP was present in the additional dog breeds tested, hereditary ataxia has been reported in Labrador Retrievers [63, 64] but not in Standard Poodles, and *NSD1* is a less likely candidate gene based on published information on the gene make the SNP less likely to be causative for hereditary ataxia and genotyping was continued for *RAB24* and not *NSD1*.

The *RAB24* polymorphism was a change from an A to C nucleotide and was located at position 113 in the first of the eight exons. It results in an amino acid change from glutamine (Q) to proline (P) at position 38 (**Figure 1**).

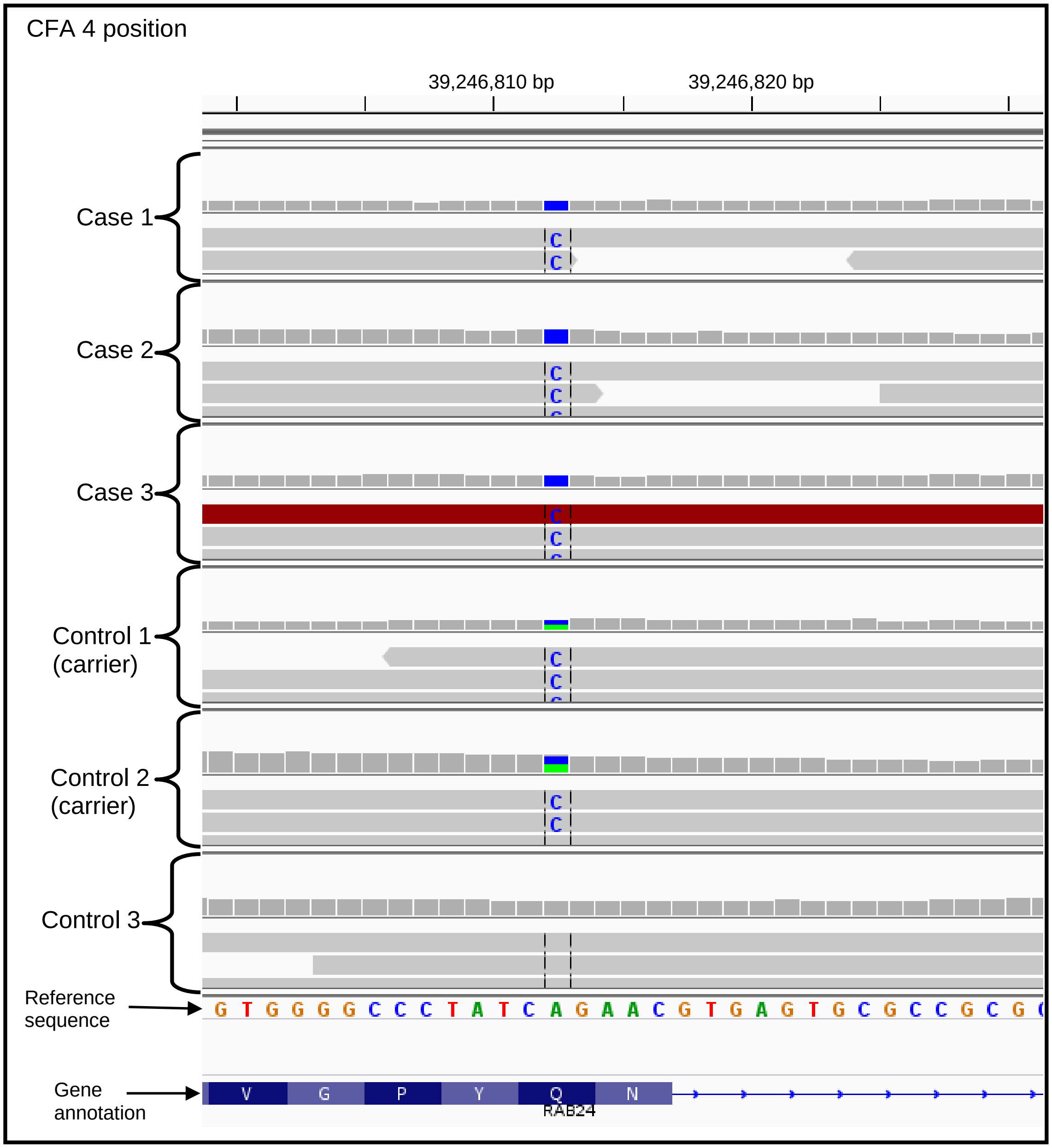


Figure 1: *RAB24* non-synonymous SNP in the six sequenced Old English Sheepdogs viewed in IGV.

In order to investigate whether the Rab24 Q38P could be causing hereditary ataxia in Old English Sheepdogs, a total of 376 Old English Sheepdogs were genotyped, all of which came from case blood lines, including 20 cases. Of these dogs, all 20 confirmed cases were homozygous for the alternate allele. 109 controls were heterozygotes, four controls were homozygous for the alternate allele, while the remainder were homozygous for the wild type allele. This cohort of dogs had been sampled approximately 20 years previously and at the time, their owners provided the dogs' phenotype. Of the four dogs that were reported as normal but exhibited the case genotype, two were littermates of affected dogs with one confirmed affected parent and one parent confirmed as a carrier. The remaining two dogs were descended from parents who were carriers. Physical examinations were not performed on these dogs by a veterinarian, so their phenotype could not be confirmed.

An additional 254 DNA samples from Old English Sheepdogs were obtained from the OFFA CHIC DNA repository. Detailed phenotypic information was not available on these dogs; although, none were reported to have hereditary ataxia on health questionnaires completed by owners at the time of DNA submission. This cohort of dogs came from approximately 70 different breeding kennels. These samples were genotyped in order to determine the prevalence of the mutation in a wider population of dogs. Twenty-eight of these dogs were heterozygous, while the rest were homozygous for the reference allele, giving an overall alternate allele prevalence in our original pedigrees and the random selection of dogs (a total of 630 dogs) of 14.3%.

The effect of the Rab24 Q38P change on protein function was investigated using Polyphen-2, a homology-based tool used to predict the impact of non-synonymous amino acid changes [65]. The glutamine (Q) to proline (P) amino acid change was predicted to probably damage protein function, with a score of 0.989 (sensitivity: 0.72 and specificity: 0.97). Using NCBI genome resources, the protein was determined to be highly conserved in other species with the glutamine amino acid being conserved in seven vertebrate species (**Figure 2**).

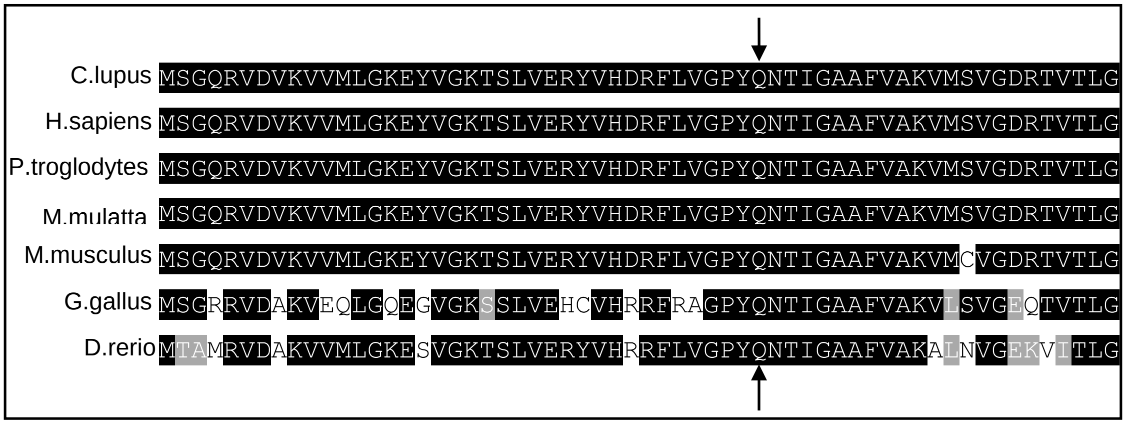


Figure 2: Amino acids 1-60 of Rab24 in seven vertebrate species aligned using ClustalW [66] and shaded with BOXSHADE 3.21. Black shading indicates identical amino acids, gray shading similar amino acids and white shading non-identical or non-similar amino acids. Arrows depict the conservation of glutamine at position 38.

Hereditary ataxia has been reported in numerous dog breeds. In order to determine whether the Q38P mutation was found in affected dogs from other breeds, DNA samples from a Dalmatian, a Beagle (both confirmed by neurological evaluation and the presence of cerebellar atrophy on MRI of the brain), 2 Rhodesian Ridgebacks (necropsy confirmed), 2 Gordon Setters (one of which was necropsy confirmed and the other confirmed by neurological evaluation and the presence of cerebellar atrophy on MRI of the brain), 2 Scottish Terriers (one MRI and one necropsy confirmed) and a Labrador retriever (diagnosed by neurological evaluation and clinical history only) were genotyped for the *RAB24* and the *NSD1* SNPs. In addition, all exons of the *RAB24* gene were sequenced in these affected dogs. The alternate allele for *RAB24 or NSD1* was not present in any of the breeds except the affected Gordon Setters. Two heterozygous synonymous SNPs were identified in exon 3 of *RAB24* in the Beagle which were also present in the Old English Sheepdog next generation sequencing data. Heterozygous non-exonic SNPs were also identified in the Dalmatian. None of these SNPs were consistent with the mode of inheritance or pathologic in nature.

Hereditary ataxia in Gordon Setters has been described in detail [22, 67] and is identical to the description of Old English Sheepdogs in terms of age of onset and progression of signs. A total of 18 affected Gordon Setters were genotyped, all of which were homozygous for the mutation. DNA from ten of these cases was obtained from archived tissue from a research dog colony [22]. An additional 82 normal Gordon Setters were genotyped from Scandinavia and the US; 24 dogs were heterozygotes and 58 were homozygous for the wild type allele. None were homozygous for the alternate allele. Excluding the 10 cases that were obtained from archival material and therefore not part of the general breeding population, there was an alternate allele frequency of 22.2%. To compare the genetic background of Gordon Setters and Old English Sheepdogs further, a cohort of seven affected and 26 control Gordon Setters were genotyped on the Illumina CanineHD BeadChip, and the genotypes on chromosome 4 were compared to those of the Old English Sheepdogs. This revealed an identical region of homozygosity extending from 39,245,536bp to 41,172,873bp in both breeds of dog (**Figure 3**).

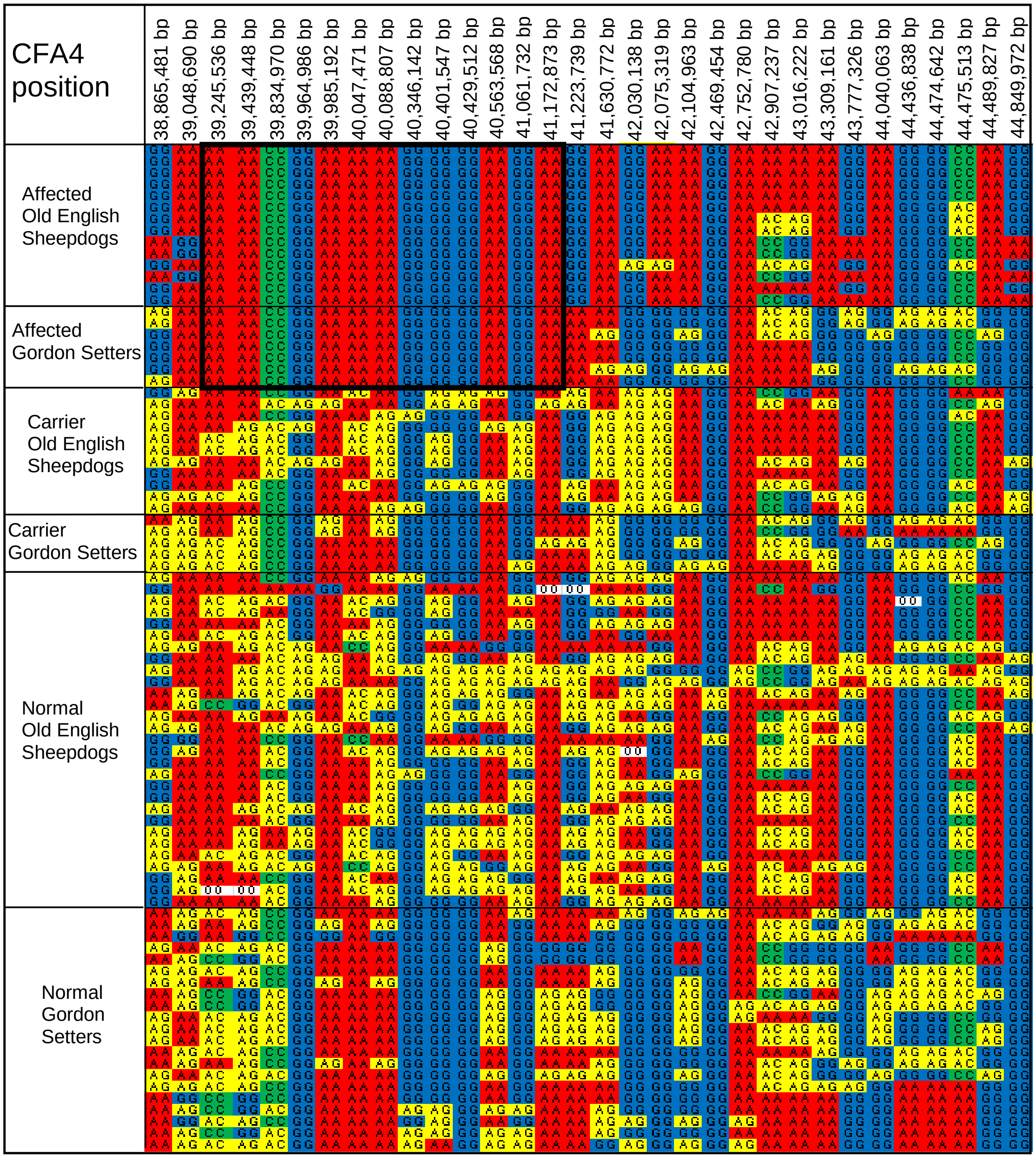


Figure 3: Genotypes of Old English Sheepdogs and Gordon Setters genotyped on Illumina arrays for a region of chromosome 4 (CFA4). Heterozygous genotypes are shaded yellow, homozygous genotypes are blue, red, and green depending on the nucleotide. White boxes indicates missing genotype. The solid black box is highlighting the shared homozygous region in affected Old English Sheepdogs and Gordon Setters.

The mean sequence coverage of the six Old English Sheepdogs in this region of shared homozygosity was 47X, making it unlikely that additional segregating variants were missed. These findings are highly supportive of the hypothesis that the *RAB24* mutation is the causative mutation of hereditary ataxia in these two breeds of dog.

In order to determine whether the *RAB24* mutation was present in other breeds of dog representing diverse ancestral lineage, DNA samples were collected from at least eight individuals from breeds in each of ten breed clusters reported as related ancestrally [68, 69]. A total of 194 individuals from 43 different breeds were genotyped (**Table 6**). All additional breeds were homozygous for the wild type allele on the *RAB24* SNP.

Table 6: Additional dog breeds genotyped on *RAB24* and *NSD1* SNPs. None of the dogs had hereditary ataxia.

	Breed	No. dogs genotyped on *Rab24* SNP	No. dogs genotyped on *NSD1* SNP
1	Airdale Terrier	1	
2	Akita	1	
3	American Staffordshire Terrier	8	7
4	Australian Shepherd	9	
5	Beagle	3	
6	Berger Picard	1	
7	Border Collie	8	
8	Boston Terrier	3	
9	Boxer	8	7
10	Brittany Spaniel	4	
11	Cairn Terrier	4	
12	Chihuahua	5	
13	Chow	1	
14	Cocker Spaniel	7	
15	Dachshund	8	
16	Doberman Pinscher	8	6
17	Dogue De Bordeaux	1	

Table 6: Continued

18	English Cocker Spaniel	1	
19	English Pointer	1	
20	English Setter	3	
21	English Springer Spaniel	1	
22	Golden Retriever	5	
23	Greyhound	6	
24	German Shepherd	11	10
25	Irish Setter	9	
26	King Charles Spaniel	14	
27	Labrador Retriever	11	7
28	Mastiff	1	
29	Miniature Schnauzer	1	
30	Pekingese	3	
31	Pembroke Welsh Corgi	2	
32	Pug	1	
33	Scottish Deerhound	7	
34	Scottish Terrier	4	4
35	Shetland Sheepdog	1	
36	Shiba Inu	1	
37	Shih Tzu	3	
38	Siberian Husky	4	
39	Spitz	2	
40	Standard Poodle	8	7
41	Weimaraner	2	
42	West Highland White Terrier	11	10
43	Wheaton Terrier	1	

RNA Sequencing and qRT-PCR

Expression levels of *RAB24* in the cerebellum were compared between case (n=4) and control (n=2) dogs by qRT-PCR. There was no significant difference in level of *RAB24* expression between cases and controls (p-value = 0.71). The A to C change in the *RAB24* transcript was present in all affected dogs and absent in the control dogs. The predicted exon/intron boundaries were also confirmed.

Discussion

Targeted next generation sequencing and subsequent genotyping identified a missense mutation that segregates with hereditary ataxia in the Old English Sheepdog and Gordon Setter dog breeds. The mutation is located in the first exon of the *RAB24* gene and changes the amino acid at position 38 from a glutamine to a proline. Rab24 belongs to the family of Ras-related small GTPases. Currently, over 75 Rab genes have been identified in vertebrates [70]. Rab (ras analog in brain) GTPases are involved in regulation of vesicular protein transport [71]. Rabs function as molecular switches, alternating between an active GTP-bound and inactive GDP-bound state accomplished by interacting with other effector molecules for GTP loading, GTP hydrolysis, inhibition of GTPase activity, and inhibition of GDP dissociation [71]. Rab proteins along with their effectors and SNARES coordinate the steps in the exocytic and endocytic pathways [72]. Five specific Rab protein features have been identified near the putative switch I region and switch II regions; regions which are unique for the GTP and GDP bound conformation of the protein [73, 74]. Rab proteins are typically prenylated at the carboxy terminus with two cysteine residues which serves to anchor the protein in membranes [73, 75].

Rab24 was first identified by PCR cloning using conserved sequences of Ras and Ras-related proteins [76]. The protein was found to be ubiquitously expressed in mouse tissues with the highest levels detected in the brain [77]. The protein localizes to the ER-cis-Golgi membranes and to late endosomal structures [77, 78]; however, when human Rab24 was expressed in COS-7 cells, the protein predominately localized to the nucleus [79]. Rab24 has characteristics different from most other Rab GTPases, such as decreased GTPase activity and reduced prenylation [80]. Rab24 binds to the N-terminal domain of SNAP29, a ubiquitous synaptosomal-associated SNARE protein, in a GTP-independent manner [81].

While no disease causing mutations have been reported in Rab24, mutations in additional Rab family members have been shown to cause disease in humans.

Mutations in Rab7 are responsible for Charcot-Marie-Tooth type 2B (CMT2B) disease, a peripheral neuropathy characterized by sensory loss and muscle weakness of the lower leg and foot [82, 83]. Rab 7 is involved in the transport between endosomes and lysosomes and has a role in microtubule transport via dynein-dynactin motors [82]. Mutations in Rab27A cause Griscelli syndrome type 2, a haemophagocytic syndrome which presents with partial albinism [84, 85]. While many of the reported mutations in Rab27A causing Griscelli type 2 result in a non-functional or absent protein, a missense mutation was identified in the switch I region demonstrating the importance of the region in effector binding and function [86]. The missense mutation identified in Rab24 of Old English Sheepdogs and Gordon Setters is located in the putative switch I region which interacts with GTP and the Mg+2 ion and contains the effector binding loop [87]. Expression of the protein in the cerebellum was not found to be reduced in affected dogs; therefore, the mutation's mechanism of action may involve reduced or absent interaction with its effector protein.

While defects in Rab proteins (Rab7) can cause neurodegeneration, mutations in additional cellular trafficking proteins can also result in neurodegeneration. *ALS2*, a guanine nucleotide exchange factor (GEF) of Rab5 involved in endosomal trafficking, has mutations in the protein resulting in juvenile autosomal recessive forms of motor neuron disease [88]. Mutations in the β-III spectrin gene are responsible for spinocerebellar ataxia type 5 disorder in humans as well as in a neonatal cortical cerebellar degeneration in the Beagle [14, 33]. The protein has multiple roles including vesicle trafficking from the Golgi and interaction with the Arp1 unit of the dynein-dynactin complex [89].

Rab24 appears to be involved in the process of autophagy [71]. The cellular localization of Rab24 to ER-cis-Golgi membranes and to late endosomal structures suggests its involvement in degradation of misfolded proteins in the ER [77]. Rab24 overexpressed in CHO cells showed a change in cellular distribution upon starvation and co-localized with L3, a marker for autophagosomes [78]. When Rab24 was mutated in part of the sequence known to be involved in guanine nucleotide binding, the protein localized to cellular inclusions, similar to inclusions seen in polyglutamine disorders, instead of the typical diffuse reticular staining seen when the wild-type protein is expressed in cell cultures [90]. In a yeast-two hybrid system, Rab24 co-localized with GABARAP (a human LC3 homolog) and cyclophilin A, suggesting its involvement in neurodegenerative disorders in addition to autophagy [79]. Autophagy has been implicated in neurodegenerative disorders as it is involved in the degradation and removal of damaged or aggregated proteins [91].

Alzheimer's disease is a neurodegenerative disorder of which autophagy plays a role in the pathogenesis. Autophagic vacuoles, rare in neurons of normal adult brain, are seen in dystrophic neurons of brains of Alzheimer's disease brains and in a mouse model of Alzheimer's disease implicating a defect in the autophagy process [92, 93]. In addition, a mouse model of Alzheimer's disease in which there is a heterozygous deletion in the gene for beclin 1, a pro-autophagic protein, suggests defects in autophagy can cause neurodegeneration [94]. Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative motor neuron disease in which mutations in the *SOD1* gene have been identified in familial cases [95]. Studies using a transgenic *SOD1* mouse model of ALS identified increases in autophagy [96] and increased autophagosomes [97]. Huntington's disease is a movement disorder resulting from the accumulation of abnormal huntingtin protein aggregates in the brain. An in vitro model of Huntington's disease showed autophagy involvement by the accumulation of mutant huntingtin in autophagosomes [98]. Finally, while autophagy and its defects have been shown to be involved in Alzheimer's disease, ALS, and Huntington's disease, more direct evidence of autophagy defects causing neurodegneration have been shown in mice. Two autophagy genes Atg5 and Atg7 have been knocked out individually in the central nervous system of mice with both cases resulting in neurodegeneration in the mice [99, 100].

The mapping of an identical mutation in the *RAB24* gene in two diverse dog breeds with a characteristically similar hereditary ataxia along with its involvement in autophagy, a pathway that when disturbed is shown to result in neurodegeneration, strongly implies that the identified mutation is the cause of hereditary ataxia in the Old English Sheepdog and Gordon Setter breeds. Further studies into how the specific mutation affects the protein function are necessary. The genetic work reported in these dissertation studies along with additional genetic linkage results and tissue examinations including: histopathology, immunohistochemistry, and ultrastuctural assessment in these breeds has been submitted for publication to PNAS and are currently under review.

References

1. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors.** Proc Natl Acad Sci U S A 1977, **74**(12):5463-5467.

2. Stranneheim H, Lundeberg J: **Stepping stones in DNA sequencing.** Biotechnol J 2012, **7**(9):1063-1073.

3. Glenn TC: **Field guide to next-generation DNA sequencers.** Mol Ecol Resour 2011, **11**(5):759-769.

4. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA: **Direct selection of human genomic loci by microarray hybridization.** Nat Methods 2007, **4**(11):903-905.

5. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR: **Genome-wide in situ exon capture for selective resequencing.** Nat Genet 2007, **39**(12):1522-1527.

6. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME: **Microarray-based genomic selection for high-throughput resequencing.** Nat Methods 2007, **4**(11):907-909.

7. Teer JK, Bonnycastle LL, Chines PS, Hansen NF, Aoyama N, Swift AJ, Abaan HO, Albert TJ, NISC Comparative Sequencing Program, Margulies EH, Green ED,

Collins FS, Mullikin JC, Biesecker LG: **Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing.** Genome Res 2010, **20**(10):1420-1431.

8. Ewing B, Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities.** Genome Res 1998, **8**(3):186-194.

9. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** Bioinformatics 2009, **25**(14):1754-1760.

10. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: **The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.** Genome Res 2010, **20**(9):1297-1303.

11. Forman OP, Penderis J, Hartley C, Hayward LJ, Ricketts SL, Mellersh CS: **Parallel mapping and simultaneous sequencing reveals deletions in BCAN and FAM83H associated with discrete inherited disorders in a domestic dog breed.** PLoS Genet 2012, **8**(1):e1002462.

12. Hytonen MK, Arumilli M, Lappalainen AK, Kallio H, Snellman M, Sainio K, Lohi H: **A novel GUSB mutation in Brazilian terriers with severe skeletal abnormalities defines the disease as mucopolysaccharidosis VII.** PLoS One 2012, **7**(7):e40281.

13. Olsson M, Meadows JR, Truve K, Rosengren Pielberg G, Puppo F, Mauceli E, Quilez J, Tonomura N, Zanna G, Docampo MJ, Bassols A, Avery AC, Karlsson EK, Thomas A, Kastner DL, Bongcam-Rudloff E, Webster MT, Sanchez A, Hedhammar A, Remmers EF, Andersson L, Ferrer L, Tintle L, Lindblad-Toh K: **A novel unstable duplication upstream of HAS2 predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs.** PLoS Genet 2011, **7**(3):e1001332.

14. Forman OP, De Risio L, Stewart J, Mellersh CS, Beltran E: **Genome-wide mRNA sequencing of a single canine cerebellar cortical degeneration case leads to the identification of a disease associated SPTBN2 mutation.** BMC Genet 2012, **13**(1):55.

15. Siso S, Navarro C, Hanzlicek D, Vandevelde M: **Adult onset thalamocerebellar degeneration in dogs associated to neuronal storage of ceroid lipopigment.** Acta Neuropathol 2004, **108**(5):386-392.

16. Summers JF, Diesel G, Asher L, McGreevy PD, Collins LM: **Inherited defects in pedigree dogs. Part 2: Disorders that are not related to breed standards.** Vet J 2010, **183**(1):39-45.

17. de Lahunta A: **Abiotrophy in domestic animals: a review.** Can J Vet Res 1990, **54**(1):65-76.

18. Siso S, Hanzlicek D, Fluehmann G, Kathmann I, Tomek A, Papa V, Vandevelde M: **Neurodegenerative diseases in domestic animals: a comparative review.** Vet J 2006, **171**(1):20-38.

19. Thames RA, Robertson ID, Flegel T, Henke D, O'Brien DP, Coates JR, Olby NJ: **Development of a morphometric magnetic resonance image parameter suitable for distinguishing between normal dogs and dogs with cerebellar atrophy.** Vet Radiol Ultrasound 2010, **51**(3):246-253.

20. Steinberg HS, Van Winkle T, Bell JS, de Lahunta A: **Cerebellar degeneration in Old English Sheepdogs.** J Am Vet Med Assoc 2000, **217**(8):1162-1165.

21. Jokinen TS, Rusbridge C, Steffen F, Viitmaa R, Syrja P, De Lahunta A, Snellman M, Cizinauskas S: **Cerebellar cortical abiotrophy in Lagotto Romagnolo dogs.** J Small Anim Pract 2007, **48**(8):470-473.

22. Cork LC, Troncoso JC, Price DL: **Canine inherited ataxia.** Ann Neurol 1981, **9**(5):492-498.

23. de Lahunta A, Fenner WR, Indrieri RJ, Mellick PW, Gardner S, Bell JS: **Hereditary cerebellar cortical abiotrophy in the Gordon Setter.** J Am Vet Med Assoc 1980, **177**(6):538-541.

24. Higgins RJ, LeCouteur RA, Kornegay JN, Coates JR: **Late-onset progressive spinocerebellar degeneration in Brittany Spaniel dogs.** Acta Neuropathol 1998, **96**(1):97-101.

25. Tatalick LM, Marks SL, Baszler TV: **Cerebellar abiotrophy characterized by granular cell loss in a Brittany.** Vet Pathol 1993, **30**(4):385-388.

26. Chrisman CL, Spencer CP, Crane SW, Mayhew IG, Averill DR, Buergelt CD: **Late-onset cerebellar degeneration in a dog.** J Am Vet Med Assoc 1983, **182**(7):717-720.

27. Gumber S, Cho DY, Morgan TW: **Late onset of cerebellar abiotrophy in a boxer dog.** Vet Med Int 2010, **2010**:406275.

28. Urkasemsin G, Linder KE, Bell JS, de Lahunta A, Olby NJ: **Hereditary cerebellar degeneration in Scottish terriers.** J Vet Intern Med 2010, **24**(3):565-570.

29. deLahunta A, Averill DR,Jr: **Hereditary cerebellar cortical and extrapyramidal nuclear abiotrophy in Kerry Blue Terriers.** J Am Vet Med Assoc 1976, **168**(12):1119-1124.

30. Coates JR, O'Brien DP, Kline KL, Storts RW, Johnson GC, Shelton GD, Patterson EE, Abbott LC: **Neonatal cerebellar ataxia in Coton de Tulear dogs.** J Vet Intern Med 2002, **16**(6):680-689.

31. Zeng R, Farias FH, Johnson GS, McKay SD, Schnabel RD, Decker JE, Taylor JF, Mann CS, Katz ML, Johnson GC, Coates JR, O'Brien DP: **A truncated retrotransposon disrupts the GRM1 coding sequence in Coton de Tulear dogs with Bandera's neonatal ataxia.** J Vet Intern Med 2011, **25**(2):267-272.

32. Kyostila K, Cizinauskas S, Seppala EH, Suhonen E, Jeserevics J, Sukura A, Syrja P, Lohi H: **A SEL1L Mutation Links a Canine Progressive Early-Onset Cerebellar Ataxia to the Endoplasmic Reticulum-Associated Protein Degradation (ERAD) Machinery.** PLoS Genet 2012, **8**(6):e1002759.

33. Ikeda Y, Dick KA, Weatherspoon MR, Gincel D, Armbrust KR, Dalton JC, Stevanin G, Durr A, Zuhlke C, Burk K, Clark HB, Brice A, Rothstein JD, Schut LJ, Day JW, Ranum LP: **Spectrin mutations cause spinocerebellar ataxia type 5.** Nat Genet 2006, **38**(2):184-190.

34. Manto M, Marmolino D: **Animal models of human cerebellar ataxias: a cornerstone for the therapies of the twenty-first century.** Cerebellum 2009, **8**(3):137-154.

35. Klockgether T, Paulson H: **Milestones in ataxia.** Mov Disord 2011, **26**(6):1134-1141.

36. Teive HA: **Spinocerebellar ataxias.** Arq Neuropsiquiatr 2009, **67**(4):1133-1142.

37. Duenas AM, Goold R, Giunti P: **Molecular pathogenesis of spinocerebellar ataxias.** Brain 2006, **129**(Pt 6):1357-1370.

38. Iwata A, Christianson JC, Bucci M, Ellerby LM, Nukina N, Forno LS, Kopito RR: **Increased susceptibility of cytoplasmic over nuclear polyglutamine aggregates to autophagic degradation.** Proc Natl Acad Sci U S A 2005, **102**(37):13135-13140.

39. Waters MF, Minassian NA, Stevanin G, Figueroa KP, Bannister JP, Nolte D, Mock AF, Evidente VG, Fee DB, Muller U, Durr A, Brice A, Papazian DM, Pulst SM: **Mutations in voltage-gated potassium channel KCNC3 cause degenerative and developmental central nervous system phenotypes.** Nat Genet 2006, **38**(4):447-451.

40. Yabe I, Sasaki H, Chen DH, Raskind WH, Bird TD, Yamashita I, Tsuji S, Kikuchi S, Tashiro K: **Spinocerebellar ataxia type 14 caused by a mutation in protein kinase C gamma.** Arch Neurol 2003, **60**(12):1749-1751.

41. van Swieten JC, Brusse E, de Graaf BM, Krieger E, van de Graaf R, de Koning I, Maat-Kievit A, Leegwater P, Dooijes D, Oostra BA, Heutink P: **A mutation in the fibroblast growth factor 14 gene is associated with autosomal dominant cerebellar ataxia [corrected.** Am J Hum Genet 2003, **72**(1):191-199.

42. Pandolfo M: **Friedreich ataxia: the clinical picture.** J Neurol 2009, **256 Suppl 1**:3-8.

43. Koeppen AH, Mazurkiewicz JE: **Friedreich ataxia: neuropathology revised.** J Neuropathol Exp Neurol 2013, **72**(2):78-90.

44. Sharma A, Buxi G, Yadav R, Kohli A: **Ataxia telangiectasia: A report of two cousins and review of literature.** Indian J Med Paediatr Oncol 2011, **32**(4):217-222.

45. Savitsky K, Bar-Shira A, Gilad S, Rotman G, Ziv Y, Vanagaite L, Tagle DA, Smith S, Uziel T, Sfez S, Ashkenazi M, Pecker I, Frydman M, Harnik R, Patanjali SR, Simmons A, Clines GA, Sartiel A, Gatti RA, Chessa L, Sanal O, Lavin MF, Jaspers NG, Taylor AM, Arlett CF, Miki T, Weissman SM, Lovett M, Collins FS, Shiloh Y: **A single ataxia telangiectasia gene with a product similar to PI-3 kinase.** Science 1995, **268**(5218):1749-1753.

46. Shiloh Y, Ziv Y: **The ATM protein: the importance of being active.** J Cell Biol 2012, **198**(3):273-275.

47. Sato J, Sasaki S, Yamada N, Tsuchitani M: **Hereditary cerebellar degenerative disease (cerebellar cortical abiotrophy) in rabbits.** Vet Pathol 2012, **49**(4):621-628.

48. Zuo J, De Jager PL, Takahashi KA, Jiang W, Linden DJ, Heintz N: **Neurodegeneration in Lurcher mice caused by mutation in delta2 glutamate receptor gene.** Nature 1997, **388**(6644):769-773.

49. Lalouette A, Guenet JL, Vriz S: **Hotfoot mouse mutations affect the delta 2 glutamate receptor gene and are allelic to lurcher.** Genomics 1998, **50**(1):9-13.

50. Lalonde R, Strazielle C: **Spontaneous and induced mouse mutations with cerebellar dysfunctions: behavior and neurochemistry.** Brain Res 2007, **1140**:51-74.

51. Lalouette A, Lohof A, Sotelo C, Guenet J, Mariani J: **Neurobiological effects of a null mutation depend on genetic context: comparison between two hotfoot alleles of the delta-2 ionotropic glutamate receptor.** Neuroscience 2001, **105**(2):443-455.

52. Fernandez-Gonzalez A, La Spada AR, Treadaway J, Higdon JC, Harris BS, Sidman RL, Morgan JI, Zuo J: **Purkinje cell degeneration (pcd) phenotypes caused by mutations in the axotomy-induced gene, Nna1.** Science 2002, **295**(5561):1904-1906.

53. Akita K, Arai S, Ohta T, Hanaya T, Fukuda S: **Suppressed Nna1 gene expression in the brain of ataxic Syrian hamsters.** J Neurogenet 2007, **21**(1-2):19-29.

54. Hamilton BA, Frankel WN, Kerrebrock AW, Hawkins TL, FitzHugh W, Kusumi K, Russell LB, Mueller KL, van Berkel V, Birren BW, Kruglyak L, Lander ES: **Disruption of the nuclear hormone receptor RORalpha] in staggerer mice.** Nature 1996, **379**(6567):736-739.

55. D'Arcangelo G, Miao GG, Chen SC, Soares HD, Morgan JI, Curran T: **A protein related to extracellular matrix proteins deleted in the mouse mutant reeler.** Nature 1995, **374**(6524):719-723.

56. Brault LS, Cooper CA, Famula TR, Murray JD, Penedo MC: **Mapping of equine cerebellar abiotrophy to ECA2 and identification of a potential causative mutation affecting expression of MUTYH.** Genomics 2011, **97**(2):121-129.

57. Ning Z, Cox AJ, Mullikin JC: **SSAHA: a fast search method for large DNA databases.** Genome Res 2001, **11**(10):1725-1729.

58. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: **Integrative genomics viewer.** Nat Biotechnol 2011, **29**(1):24-26.

59. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** Methods Mol Biol 2000, **132**:365-386.

60. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL: **Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction.** BMC Bioinformatics 2012, **13**:134-2105-13-134.

61. Marshall OJ: **PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR.** Bioinformatics 2004, **20**(15):2471-2472.

62. Brinkhof B, Spee B, Rothuizen J, Penning LC: **Development and evaluation of canine reference genes for accurate quantification of gene expression.** Anal Biochem 2006, **356**(1):36-43.

63. Perille AL, Baer K, Joseph RJ, Carrillo JM, Averill DR: **Postnatal cerebellar cortical degeneration in Labrador Retriever puppies.** Can Vet J 1991, **32**(10):619-621.

64. Bildfell RJ, Mitchell SK, de Lahunta A: **Cerebellar cortical degeneration in a Labrador retriever.** Can Vet J 1995, **36**(9):570-572.

65. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR: **A method and server for predicting damaging missense mutations.** Nat Methods 2010, **7**(4):248-249.

66. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** Nucleic Acids Res 1994, **22**(22):4673-4680.

67. Steinberg HS, Troncoso JC, Cork LC, Price DL: **Clinical features of inherited cerebellar degeneration in Gordon setters.** J Am Vet Med Assoc 1981, **179**(9):886-890.

68. Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, Degenhardt JD, Boyko AR, Earl DA, Auton A, Reynolds A, Bryc K, Brisbin A, Knowles JC, Mosher DS, Spady TC, Elkahloun A, Geffen E, Pilot M, Jedrzejewski W, Greco C, Randi E, Bannasch D, Wilton A, Shearman J, Musiani M, Cargill M, Jones PG, Qian Z, Huang W, Ding ZL, Zhang YP, Bustamante CD, Ostrander EA, Novembre J, Wayne RK: **Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication.** Nature 2010, **464**(7290):898-902.

69. Parker HG: **Genomic analyses of modern dog breeds.** Mamm Genome 2012, **23**(1-2):19-27.

70. Harris KP, Littleton JT: **Vesicle trafficking: a Rab family profile.** Curr Biol 2011, **21**(20):R841-3.

71. Barnekow A, Thyrock A, Kessler D: **Chapter 5: rab proteins and their interaction partners.** Int Rev Cell Mol Biol 2009, **274**:235-274.

72. Zerial M, McBride H: **Rab proteins as membrane organizers.** Nat Rev Mol Cell Biol 2001, **2**(2):107-117.

73. Pereira-Leal JB, Seabra MC: **The mammalian Rab family of small GTPases: definition of family and subfamily sequence motifs suggests a mechanism for functional specificity in the Ras superfamily.** J Mol Biol 2000, **301**(4):1077-1087.

74. Pfeffer SR: **Structural clues to Rab GTPase functional diversity.** J Biol Chem 2005, **280**(16):15485-15488.

75. Chavrier P, Goud B: **The role of ARF and Rab GTPases in membrane transport.** Curr Opin Cell Biol 1999, **11**(4):466-475.

76. Chavrier P, Simons K, Zerial M: **The complexity of the Rab and Rho GTP-binding protein subfamilies revealed by a PCR cloning approach.** Gene 1992, **112**(2):261-264.

77. Olkkonen VM, Dupree P, Killisch I, Lutcke A, Zerial M, Simons K: **Molecular cloning and subcellular localization of three GTP-binding proteins of the rab subfamily.** J Cell Sci 1993, **106 (Pt 4)**(Pt 4):1249-1261.

78. Munafo DB, Colombo MI: **Induction of autophagy causes dramatic changes in the subcellular distribution of GFP-Rab24.** Traffic 2002, **3**(7):472-482.

79. Wu M, Yin G, Zhao X, Ji C, Gu S, Tang R, Dong H, Xie Y, Mao Y: **Human RAB24, interestingly and predominantly distributed in the nuclei of COS-7 cells, is colocalized with cyclophilin A and GABARAP.** Int J Mol Med 2006, **17**(5):749-754.

80. Erdman RA, Shellenberger KE, Overmeyer JH, Maltese WA: **Rab24 is an atypical member of the Rab GTPase family. Deficient GTPase activity, GDP dissociation inhibitor interaction, and prenylation of Rab24 expressed in cultured cells.** J Biol Chem 2000, **275**(6):3848-3856.

81. Schardt A, Brinkmann BG, Mitkovski M, Sereda MW, Werner HB, Nave KA: **The SNARE protein SNAP-29 interacts with the GTPase Rab3A: Implications for**

membrane trafficking in myelinating glia. J Neurosci Res 2009, **87**(15):3465-3479.

82. Verhoeven K, De Jonghe P, Coen K, Verpoorten N, Auer-Grumbach M, Kwon JM, FitzPatrick D, Schmedding E, De Vriendt E, Jacobs A, Van Gerwen V, Wagner K, Hartung HP, Timmerman V: **Mutations in the small GTP-ase late endosomal protein RAB7 cause Charcot-Marie-Tooth type 2B neuropathy.** Am J Hum Genet 2003, **72**(3):722-727.

83. Bucci C, De Luca M: **Molecular basis of Charcot-Marie-Tooth type 2B disease.** Biochem Soc Trans 2012, **40**(6):1368-1372.

84. Menasche G, Pastural E, Feldmann J, Certain S, Ersoy F, Dupuis S, Wulffraat N, Bianchi D, Fischer A, Le Deist F, de Saint Basile G: **Mutations in RAB27A cause Griscelli syndrome associated with haemophagocytic syndrome.** Nat Genet 2000, **25**(2):173-176.

85. Meeths M, Bryceson YT, Rudd E, Zheng C, Wood SM, Ramme K, Beutel K, Hasle H, Heilmann C, Hultenby K, Ljunggren HG, Fadeel B, Nordenskjold M, Henter JI: **Clinical presentation of Griscelli syndrome type 2 and spectrum of RAB27A mutations.** Pediatr Blood Cancer 2010, **54**(4):563-572.

86. Westbroek W, Tuchman M, Tinloy B, De Wever O, Vilboux T, Hertz JM, Hasle H, Heilmann C, Helip-Wooley A, Kleta R, Gahl WA: **A novel missense mutation (G43S) in the switch I region of Rab27A causing Griscelli syndrome.** Mol Genet Metab 2008, **94**(2):248-254.

87. Dumas JJ, Zhu Z, Connolly JL, Lambright DG: **Structural basis of activation and GTP hydrolysis in Rab proteins.** Structure 1999, **7**(4):413-423.

88. Hadano S, Otomo A, Kunita R, Suzuki-Utsunomiya K, Akatsuka A, Koike M, Aoki M, Uchiyama Y, Itoyama Y, Ikeda JE: **Loss of ALS2/Alsin exacerbates motor dysfunction in a SOD1-expressing mouse ALS model by disturbing endolysosomal trafficking.** PLoS One 2010, **5**(3):e9805.

89. Clarkson YL, Gillespie T, Perkins EM, Lyndon AR, Jackson M: **Beta-III spectrin mutation L253P associated with spinocerebellar ataxia type 5 interferes with binding to Arp1 and protein trafficking from the Golgi.** Hum Mol Genet 2010, **19**(18):3634-3641.

90. Maltese WA, Soule G, Gunning W, Calomeni E, Alexander B: **Mutant Rab24 GTPase is targeted to nuclear inclusions.** BMC Cell Biol 2002, **3**:25.

91. Garcia-Arencibia M, Hochfeld WE, Toh PP, Rubinsztein DC: **Autophagy, a guardian against neurodegeneration.** Semin Cell Dev Biol 2010, **21**(7):691-698.

92. Nixon RA, Wegiel J, Kumar A, Yu WH, Peterhoff C, Cataldo A, Cuervo AM: **Extensive involvement of autophagy in Alzheimer disease: an immuno-electron microscopy study.** J Neuropathol Exp Neurol 2005, **64**(2):113-122.

93. Yu WH, Cuervo AM, Kumar A, Peterhoff CM, Schmidt SD, Lee JH, Mohan PS, Mercken M, Farmery MR, Tjernberg LO, Jiang Y, Duff K, Uchiyama Y, Naslund J, Mathews PM, Cataldo AM, Nixon RA: **Macroautophagy--a novel Beta-amyloid peptide-generating pathway activated in Alzheimer's disease.** J Cell Biol 2005, **171**(1):87-98.

94. Pickford F, Masliah E, Britschgi M, Lucin K, Narasimhan R, Jaeger PA, Small S, Spencer B, Rockenstein E, Levine B, Wyss-Coray T: **The autophagy-related protein beclin 1 shows reduced expression in early Alzheimer disease and regulates amyloid beta accumulation in mice.** J Clin Invest 2008, **118**(6):2190-2199.

95. Son JH, Shim JH, Kim KH, Ha JY, Han JY: **Neuronal autophagy and neurodegenerative diseases.** Exp Mol Med 2012, **44**(2):89-98.

96. Li L, Zhang X, Le W: **Altered macroautophagy in the spinal cord of SOD1 mutant mice.** Autophagy 2008, **4**(3):290-293.

97. Morimoto N, Nagai M, Ohta Y, Miyazaki K, Kurata T, Morimoto M, Murakami T, Takehisa Y, Ikeda Y, Kamiya T, Abe K: **Increased autophagy in transgenic mice with a G93A mutant SOD1 gene.** Brain Res 2007, **1167**:112-117.

98. Kegel KB, Kim M, Sapp E, McIntyre C, Castano JG, Aronin N, DiFiglia M: **Huntingtin expression stimulates endosomal-lysosomal activity, endosome tubulation, and autophagy.** J Neurosci 2000, **20**(19):7268-7278.

99. Komatsu M, Waguri S, Chiba T, Murata S, Iwata J, Tanida I, Ueno T, Koike M, Uchiyama Y, Kominami E, Tanaka K: **Loss of autophagy in the central nervous system causes neurodegeneration in mice.** Nature 2006, **441**(7095):880-884.

100. Hara T, Nakamura K, Matsui M, Yamamoto A, Nakahara Y, Suzuki-Migishima R, Yokoyama M, Mishima K, Saito I, Okano H, Mizushima N: **Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice.** Nature 2006, **441**(7095):885-889.

CHAPTER 5

Conclusions, Limitations, and Future Directions

Conclusions

Canine atopic dermatitis in WHWT was investigated using a linkage study and GWAS. Three families of WHWT and a maximum of 90 dogs were genotyped at 256 microsatellite markers across the genome. Linkage analysis was unable to identify a locus for atopic dermatitis in these dogs; however, the analysis was limited to the canine autosomes and the homozygosity of markers resulted in gaps in genome coverage. Simulations performed using the pedigree structure of genotyped dogs revealed the inability to detect a mutation of low penetrance, which could explain the lack of detection of an atopic dermatitis locus. To overcome these pitfalls, a genome-wide association study was performed in a larger cohort of WHWT. Fifty-eight of the dogs from the linkage study and an additional 128 dogs were genotyped on Illumina CanineHD BeadChips. A GWAS was performed using 96 WHWT cases and 87 WHWT controls and revealed SNPs associated with the disease; however, none passed a stringent Bonferroni correction for multiple testing. Population stratification was detected and corrected for; however, did not result in significant associations after a multiple testing correction. When considering the top most associated SNPs, none replicated previous genome-wide association studies performed in WHWT or multiple breeds; however the genes *HMGCR* and *F2RL1* were located in a detected chromosome 3 region and stand out as possible candidates for atopic dermatitis. The high prevalence of atopic dermatitis in the WHWT breed and the complex nature of the disease, including environmental effects, makes it difficult to identify normal individuals as well as to be certain of a normal phenotype.

Hereditary ataxia in the Old English Sheepdog breed was investigated using a GWAS and targeted next generation sequencing. Fifty-five dogs were genotyped on Illumina SNP20 and CanineHD BeadChips and a GWAS was performed with the

12,986 SNPs common to both SNP chips. A region of significant association on chromosome 4 was identified; however, numerous genes with no clear candidates for hereditary ataxia were present in the region. Targeted capture and next generation sequencing was performed in six OES covering a continuous 12 Mb region on chromosome 4 including the region of multiple association signals from the GWAS. The successful capture produced the coverage needed to identify six novel non-synonymous sequence variants which segregated with the disease and the expected inheritance in the dogs. Genotyping these variants in additional dogs and additional breeds along with online mutation prediction tools identified a SNP in the *RAB24* gene likely to be causative of hereditary ataxia in OES and the Gordon Setter breed.

Limitations

One of the limitations in investigating the genetics of atopic dermatitis in the West Highland White Terrier breed is the unclear mode of inheritance. A proper segregation analysis could not be performed due to the difficulty in collecting the necessary phenotype information on complete litters of dogs. A previously reported segregation analysis was not very informative due to a complex ascertainment and found an equally likely fully dominant and recessive inheritance [1]. A second limitation is the difficulty in confirming normal phenotype which also resulted in difficulty in collecting the planned sample numbers. While affected dogs of all ages were collected, a normal phenotype could not be assigned to dogs below the age of five because of the chance the dog would still develop the disease after being sampled. Atopic dermatitis is defined by the exposure to environmental allergens which lends the possibility that a normal dog may become affected once exposed to the correct instigating environmental allergen.

While a likely causative mutation was identified for cerebellar degeneration in Old English Sheepdogs, there were limitations to the research which should be acknowledged. The collection of cerebellar tissue from affected dogs to use for qRT-

PCR analysis was limited. Cerebellar tissue had been collected on only two OES and two Gordon Setters, and the tissue was known to be harvested immediately after death from only one of the Gordon Setters. Additional tissue samples from affected dogs would allow a more thorough examination of the difference in gene expression levels of Rab24 in cases and controls.

Future directions

Planned research into atopic dermatitis in the WHWT includes entire genome sequencing of two affected and two normal WHWT. This will enable inspection of the sequence of all genes and intergenic regions in the areas identified by the GWAS reported in Chapter 3. *HMGCR,* HMG CoA reductase, is a good candidate gene for atopic dermatitis due to its involvement in cholesterol synthesis and maintenance of the skin's protective barrier. While the genome sequence data of four WHWT should identify any sequence variation in the gene, it is not an extremely large gene with a transcript of 6,528 bp, and Sanger sequencing of its 20 exons can be done affordably in six dogs. The other candidate gene, *F2RL1,* could also be easily Sanger sequenced as it is approximately half the size of *HMGCR*. The additional candidate genes implicated in human atopic dermatitis not showing association by GWAS can also be explored; including the *FLG* gene. The *FLG* gene sequence can be examined for the identical SNPs previously published to have an association with atopic dermatitis in small dog breeds [2]. Additional whole genome sequencing of two Scottish Terriers not affected with atopic dermatitis is also planned. Data from these two dogs will allow the comparison of the affected WHWT to a related breed with a much lower incidence of atopic dermatitis, reducing the chance of using dogs of a normal phenotype that have just not encountered the environmental allergen triggering the disease. In addition to whole genome sequencing, the collection and banking of affected and normal WHWT DNA samples is ongoing, which allows the future genotyping of additional samples on the Illumina BeadChip, thus improving GWAS power.

Ongoing research into hereditary ataxia in the OES and Gordon Setter breeds will involve further investigation of *RAB24* and the mutation identified. While no difference in expression levels of Rab24 in cerebellar tissue collected thus far was detected, collection of additional affected samples will continue and may show a change. The investigation of gene expression levels of proteins involved in the Rab24 autophagy pathway is imminent as RNA sequence data has been collected on an affected OES and Gordon Setter along with a normal Beagle control. In addition, fibroblast cell cultures have been established from cells harvested from an affected Gordon Setter. These cultures will enable further investigations into the functional effects of the Rab24 protein on autophagy as well as potential therapy for the disease.

References

1. Roque JB, O'Leary CA, Kyaw-Tanner M, Latter M, Mason K, Shipstone M, Vogelnest L, Duffy DL: **Haplotype sharing excludes canine orthologous Filaggrin locus in atopy in West Highland White Terriers.** Anim Genet 2009, **40**(5):793.

2. Suriyaphol G, Suriyaphol P, Sarikaputi M, Theerawatanasirikul S, Sailasuta A: **Association of Filaggrin (*FLG*) Gene Polymorphism with Canine Atopic Dermatitis in Small Breed Dogs.** Thai J Vet Med 2011, **41**(4):509-510-517.