Mapping Trait Genes in Dogs

Using the Dog as a Model Organism

Izabella Baranowska Körberg

Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics
Uppsala

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2011

Mapping Trait Genes in Dogs. Using the Dog as a Model Organism

Abstract

Understanding the genetic background of a given phenotypic trait or disease has long intrigued scientists. Model organisms can be employed when it is not feasible or possible to find trait causality in human cohorts. This thesis reports the mapping of both a phenotypic trait (white spotting) and a disease (sensory ataxic neuropathy) in dogs, and in doing so highlights the potential of the dog as a model organism for mapping traits of relevance to human health and biology.

For white spotting, a two-stage mapping approach was used to identify an associated genomic region, which contained only the *microphtalmia-associated transcription factor* (*MITF*) gene. Stage 1: a genome-wide association analysis in a single breed, taking advantage of the extensive within breed linkage disequilibrium (LD) and long haplotypes, identified a discrete region of approximately 1 Mb. Stage 2: finemapping in an additional breed presenting the same phenotype, exploiting the short LD and haplotypes shared across dog breeds, was used to narrow the region to about 100 kb. We functionally evaluated two candidate polymorphisms associated with *MITF*, a SINE insertion and a length polymorphism upstream of the melanocyte-specific transcription start site of *MITF*. The data indicated that both polymorphisms affect transcription from the MITF-M promoter.

Sensory ataxic neuropathy (SAN) is a neurological disorder affecting a specific maternal lineage of Golden Retrievers. We identified a one base pair deletion in mitochondrial $tRNA^{Tyr}$ and through biochemical analyses of mitochondria and functional studies of the deletion, confirmed causality and the mitochondrial origin of SAN. This is one of the first mitochondrial disorders identified in dogs and we additionally developed a genetic test for the public to allow for the elimination of the disease from this breed.

The genetic and functional analyses of both white spotting coat colour and SAN in dogs, which are controlled by mutations of the two different genomes, clearly demonstrate the utility of the canine model and establish a new role for man's best friend.

Keywords: dog, model organism, association mapping, white spotting locus, *MITF*, mitochondrial disorder, sensory ataxic neuropathy, $tRNA^{Tyr}$

Author's address: Izabella Baranowska Körberg, SLU, Department of Animal Breeding and Genetics, P.O. Box 597, 751 24 Uppsala, Sweden
E-mail: Izabella.Baranowska@slu.se

Contents

6

List of Publications

This thesis is based on the work contained in the following papers, which are referred to by Roman numerals in the text:

I   Karlsson, E. K., **Baranowska, I.**, Wade, C. M., Salmon Hillbertz, N. H. C., Zody, M. C., Anderson, N., Biagi, T. M., Patterson, N., Rosengren Pielberg, G., Kulbokas III, E. J., Comstock, K. E., Keller, E. T., Mesirov, J. P., von Euler, H., Kämpe, O., Hedhammar, Å., Lander, E. S., Andersson, G., Andersson, L. & Lindblad-Toh, K. (2007). Efficient Mapping of Mendelian Traits in Dogs Through Genome-Wide Association. *Nature Genetics* 39(11), 1321-1328.

II **Baranowska, I.**, Hultin Jäderlund, K., Nennesmo, I., Holmqvist, E., Heidrich, N., Larsson, N.-G., Andersson, G., Wagner, E. G. H., Hedhammar, Å., Wibom, R. & Andersson, L. (2009). Sensory Ataxic Neuropathy in Golden Retriever Dogs Is Caused by a Deletion in the Mitochondrial $tRNA^{Tyr}$ Gene. *PLoS Genetics* 5(5), e1000499. doi:10.1371/journal.pgen.1000499

III **Baranowska Körberg, I.**, Zhang X., Rosengren Pielberg, G., Gustafson, U., Karlsson, E. K., Hedhammar, Å., Lindblad-Toh, K., Andersson, G. & Andersson, L. Functional Studies of Sequence Polymorphisms at the *MITF* Locus Associated with White Spotting in Dogs. (manuscript).

Papers I and II are reproduced with the permission of the publishers.

Related Work by the Author

(Not included in the thesis)

Lindqvist, C., Janczak, A. M., Nätt, D., **Baranowska, I.**, Lindqvist, N., Wichman, A., Lundeberg, J., Lindberg, J., Torjesen, P. A. & Jensen, P. (2007). Transmission of Stress-Induced Learning Impairment and Associated Brain Gene Expression from Parents to Offspring in Chickens. *PLoS ONE* 2(4), e364. doi:10.1371/journal.pone.0000364

Awano, T., Johnson, G. S., Wade, C. M., Katz, M. L., Johnson, G. C., Taylor, J. F., Perloski, M., Biagi, T., **Baranowska, I.**, Long, S., March, P. A., Olby, N. J., Shelton, G. D., Khan, S., O'Brien, D. P., Lindblad-Toh, K. & Coates, J. R. (2009). Genome-Wide Association Analysis Reveals a *SOD1* Mutation in Canine Degenerative Myelopathy That Resembles Amyotrophic Lateral Sclerosis. *Proceedings of the National Academy of Science of the Unites States of America* 106(8), 2794-2799.

Hultin Jäderlund, K., **Baranowska Körberg, I.** & Nødtvedt, A. (2011). Inherited polyneuropathy in the Leonberger Breed; Pedigree Data Indicates a Major Risk Factor on X-Chromosome. *Journal of Veterinary Internal Medicine (Accepted).*

Abbreviations

ALS	amyotrophic lateral sclerosis
ATP	adenosine-5'-triphosphate
BAC	bacterial artificial chromosome
bp	base pair
CNS	central nervous system
CNV	copy number variant
COX	cytochrome c
DNA	deoxyribonucleic acid
ECM	encephalomyopathy
FBSN	familial bilateral striatial necrosis
Gb	gigabases
GWAS	genome-wide association study
kb	kilobases
LD	linkage disequilibrium
lod-score	logarithm of odds score
LOHN	Leber hereditary optic neuropathy
Lp	length polymorphism
LpL	long Lp
LpS	short Lp
LS	Leigh's syndrome
MAF	minor allele frequency
MAPR	mitochondrial ATP production rate
Mb	megabases
MELAS	mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes
MERRF	myoclonic epilepsy with red-ragged fibres
MILS	maternally inherited Leigh's syndrome
MITF	microphtalmia-associated transcription factor

mtDNA	mitochondrial DNA
NARP	neuropathy ataxia and retinitis pigmentosa
nDNA	nuclear DNA
PCR	polymerase chain reaction
PEO	progressive external ophthalmoplegia
PPK	palmoplantar keratoderma
qOLA	quantitative oligonucleotide ligation assay
RLU	relative luciferase unit
SAN	sensory ataxic neuropathy
SDH	succinate dehydrogenase
SIDS	sudden infant death syndrome
SINE	short interspersed nucleotide element
SINEC-Cf	canine-specific SINE
SLE	systemic lupus erythematosus
SLU	Swedish University of Agricultural Sciences
SNP	single nucleotide polymorphism
tRNA	transfer ribonucleic acid
$tRNA^{Cys}$	$tRNA^{Cysteine}$
$tRNA^{Gln}$	$tRNA^{Glutamine}$
$tRNA^{Tyr}$	$tRNA^{Tyrosine}$
wt	wild type

Introduction

The recent development of nucleotide sequencing technologies and bioinformatics has revolutionized biology. The release of high quality genome sequences for a variety of species enables scientists to explore the field of genomics and genetics like never before, including the fascinating and often challenging world of genetic trait mapping. Genetic mapping is the first step in understanding the molecular etiology of a given phenotypic trait or disease. Useful tools for further genetic dissection include, for example, next-generation sequencing (Schuster, 2008) allowing for efficient mutation detection, maps of evolutionary conservation (Miller *et al.*, 2007), and also databases containing information about different aspects of cell biology, such as tissue-specific gene expression and cellular pathways. *In vivo* models, both cell culture and non-human vertebrates, are crucial to the process of understanding the biological mechanisms and defining the functional roles of the causative mutations behind phenotypic traits and diseases. The current resources, and many more within reach, provide endless possibilities for researchers. Hence, this truly is the golden age of genetics.

The two most frequently used mammalian species for genetic studies of phenotypic traits and diseases are humans and mice. However domestic animals, such as the horse, chicken, dog and pig, have also been used successfully both for mapping specific production traits but also as model organisms for human disease phenotypes (Mignot *et al.*, 1991; Van Laere *et al.*, 2003; Rubin *et al.*, 2007; Rosengren Pielberg *et al.*, 2008). Thus, domestic animals are a complement to other experimental organisms for unravelling the genetics behind simple and complex traits (Andersson, 2009).

Phenotypic traits and diseases can be divided into monogenic and complex polygenic traits. Simple Mendelian traits are controlled by a single highly

penetrant locus and show a Mendelian inheritance pattern, which is either autosomal or X-linked (recessive, dominant or co-dominant), thus the relationship between the causative allele and the trait of interest is one-to-one. In contrast, multi-factorial traits are controlled by multiple loci as well as environmental factors, and therefore the inheritance does not follow any simple Mendelian pattern. Each locus usually explains a small part of the phenotypic variation. For example, it has been estimated that height in humans has an 80 % heritability, however the 180 variants associated with height together explain merely 10 % of the genetics (Lango Allen *et al.*, 2010).

Genetic trait mapping

Genetic markers

For most of the 20th century, whole-genome linkage or association studies in humans were not feasible, mainly due to too few well-defined markers (Lander & Schork, 1994; Hirschhorn & Daly, 2005; Altshuler *et al.*, 2008). It wasn't until the 1980s, when DNA polymorphisms (tandem repeats, microsatellites or simple sequence repeats) were discovered throughout the genome, that it became possible to trace inheritance and perform whole-genome linkage studies in humans (Botstein *et al.*, 1980; Altshuler *et al.*, 2008). In 1987 the first genetic linkage map was generated with approximately 400 markers (Donis-Keller *et al.*, 1987). Ten years later it was expanded to about 5000 markers (Dib *et al.*, 1996). The next advance came with the discovery and cataloguing of large numbers of single nucleotide polymorphisms (SNPs) which has now been expanded to several million SNPs (Frazer *et al.*, 2007). Due to their high abundance in the genome, biallelic nature and also the ease of scoring, SNPs are considered advantageous in whole-genome studies.

Linkage analysis and whole-genome association mapping

Methods for studying whole-genome genotype-phenotype correlations can be divided into two main types, i.e. linkage (Figure 1A) and association analysis (Figure 1B) (Lander & Schork, 1994; Kruglyak, 1999; Peltonen *et al.*, 2000; Hirschhorn & Daly, 2005; Altshuler *et al.*, 2008). Both methods are dependent on the occurrence of recombination (meiotic cross-over). In linkage analysis, recombination events are evaluated in a family under a well-defined inheritance model. The family includes parents, which are heterozygous for the trait locus studied, enabling genetic markers to segregate and to be traced over generations. Family members presenting the

12

same phenotype are expected to share alleles at markers close to the causative mutation. Due to the markers proximity to the disease gene recombination is less likely to occur, hence the genetic markers and the causative alleles are linked. The power of linkage analysis depends on the number of informative meioses (Lander & Schork, 1994). Lod (log of odds)-score tests are used to provide evidence for linkage and estimates of recombination fractions. Positive lod-scores support linkage, negative lod-scores indicate that linkage is less likely at the recombination fraction used to calculate the lod score. Linkage analysis in humans has proven to be an effective method for mapping traits with simple Mendelian inheritance caused by highly penetrant alleles, where the relationship between phenotype and genotype is one-to-one. Disease-causing mutations, for example familial early onset forms of Alzheimer's disease and Huntington's disease (Peltonen *et al.*, 2006; Bekris *et al.*, 2010; Munoz-Sanjuan & Bates, 2011) have been mapped utilizing this strategy. Although linkage analysis proved to be a successful method to identify trait loci with simple inheritance, the approach was unsatisfactory for mapping loci underlying multifactorial traits in humans. When mapping complex traits assumptions about the genotype or the inheritance model cannot be made, and consequently the power of the analysis is decreased. Instead of using linkage for mapping complex traits, whole-genome association studies were suggested during the mid 1990s (Lander, 1996; Risch & Merikangas, 1996; Collins *et al.*, 1997).
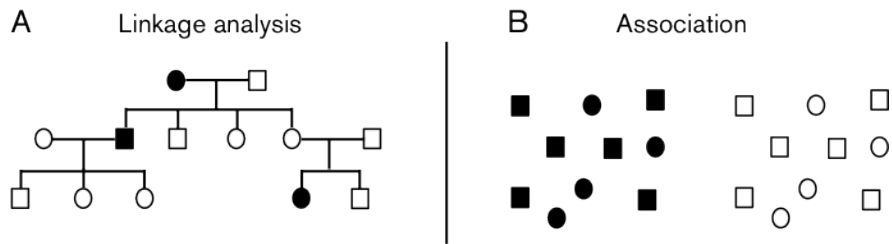
Figure 1. Schematic overview of linkage analysis and association mapping. (A) The mode of inheritance can be defined by pedigree analysis. The analysis is performed under the assumption that alleles causing a phenotypic trait or disease will cosegregate with genetic markers on the same chromosome. (B) The allele frequency is estimated for each marker in unrelated individuals and compared between cases and controls. Circles represent females and squares represent males, cases are illustrated by black symbols and controls by white (figure modified from Lander & Schork (1994)).

In association mapping, historical recombination events that have occurred in the ancestors of the genotyped individuals are utilized. A prerequisite in association mapping is that the markers are dispersed equally along the genome, without regions of gaps. In association mapping, the association between neutral polymorphisms (nowadays SNPs) and a certain trait is defined using unrelated individuals. Allele frequencies are determined in cases and compared to the allele frequency in controls. If the frequency of an allele is more common in the cases than the controls, it is considered to be associated with a certain trait. The assumption in association mapping is that at least one marker will be in linkage disequilibrium (LD) with a polymorphism affecting the trait. LD refers to the non-random association of two or more alleles at different loci, hence loci located in the vicinity of each other on the same chromosome are more likely to be in LD than loci on different chromosomes. The power of association mapping depends on the degree of LD between the genotyped marker and the mutation of interest. Markers in proximity of each other will show stronger LD, since LD decays with distance. In early 2006, technologies for large-scale genotyping of large sample sets enabled genome-wide studies. By 2007, more than 150 associations between diseases and common SNPs had been identified, although the significant loci usually only explained a minor

fraction of the phenotypic variation (Consortium, 2007; McCarthy *et al.*, 2008; Manolio *et al.*, 2009).

Utility of LD and haplotype blocks in genome-wide association studies in human populations

There are several reasons why association mapping may not identify a significant signal of association, for example the study may be underpowered with regard to sample size and marker density. Thus, a central aspect of genome-wide association studies (GWAS) is to assess the minimum amount of SNPs to include without losing power to detect genetic variation. This can be defined by assessing the average extent of LD and the size of the haplotypes along the genome (Kruglyak, 1999; Consortium, 2005). While LD is the quantitative unit for how markers are linked to each other, the haplotype defines the combination of alleles at different loci that are inherited together. As the extent of LD and size of haplotype blocks differs between different ethnic groups and varies across the genome, the number of SNPs required in a GWAS may vary depending on the sample set. It has been estimated in humans that the size of average haplotype blocks are about 11 kb in Africans and approximately 22 kb in Caucasians and Asian people, with 3-5 haplotypes per block (Gabriel *et al.*, 2002; Consortium, 2005). It has been assessed that a panel of 300,000 to 1,000,000 SNPs should be sufficient for most GWAS depending on the origin of the population studied. The International HapMap Consortium was created to generate a publicly available database of common human sequence variation possibly simplifying study design when mapping different traits and diseases in humans. Initially common genetic variation was analysed in four geographically diverse human populations (Consortium, 2005; Frazer *et al.*, 2007), which was subsequently expanded to eleven populations (Altshuler *et al.*, 2010). In a study of eleven well-described isolated populations, it was estimated that LD is considerably higher in most isolates compared to outbred populations (Service *et al.*, 2006).

Characteristics of isolated populations

There are several advantages to mapping traits in isolated populations. Due to population bottlenecks, that can sometimes be re-occurring followed by periods of rapid growth, allele frequencies may change (genetic drift). As a consequence isolated populations may have a higher prevalence of certain diseases due to the enrichment of shared risk alleles (Peltonen *et al.*, 2000). Thus, there is less genetic heterogeneity in an isolated population, many patients share the same causative mutation and its associated haplotype

block. Another positive attribute lies in the reduced noise in the data set through the shared environmental factors applied to isolated populations. These populations by definition have less migration and thus more intact families, often with good genealogical records (Peltonen *et al.*, 2000). Finally, population stratification, which can be a major issue in the general population, may be less of a problem when mapping traits in isolated populations.

Sources to false-positive associations

Population stratification by definition arises when markers have different allele frequencies in different sub-populations (Hirschhorn & Daly, 2005). This means that positive signals of association may describe the difference between the different sub-populations, rather than an association with a disease trait. Family structure or cryptic relatedness further complicates any corrections for stratification. There are different methods implemented in different softwares enabling the detection of, and correction for, stratification as reviewed by Price *et al.* (2010). A possible way to avoid stratification is to use family-based samples. Although these sample sets may present other challenges such as reduced power with regard to sample size.

False-positive associations may also arise due to missing genotyping data. Some methods have a lower success rate when genotyping heterozygous positions, thus leading to deviations from Hardy-Weinberg equilibrium (Lander & Schork, 1994; Hirschhorn & Daly, 2005). However, these types of errors are generally less likely to cause false-positive associations since it is expected that cases and controls will be equally affected and is therefore more often causing false-negative results.

An alternative explanation for unsuccessful association studies may lie with mitochondrial inheritance. In this case a trait is inherited along the maternal lineage, hence there is no true association between the disease and the genetic markers of the nuclear genome.

Genetics of mitochondrial disorders

Disease-causing polymorphisms of the mitochondrial genome

The mitochondrion is an organelle located in the cytoplasm of the cell with many crucial functions, such as to provide the cell with energy in the form of adenosine-5'-triphosphate (ATP). The mitochondrion has its own genome, referred to as the mitochondrial DNA (mtDNA). However, the mtDNA is only about 16.6 kb compared to the nuclear DNA (nDNA) which is about 2.5 Gb in the dog.

The first disease-causing mutations in the human mitochondrial genome were identified during the late 1980s (Holt *et al.*, 1988; Wallace *et al.*, 1988). Since then, more than 550 disease-causing mutations and rearrangements have been associated with disorders such as cardiomyopathies, myopathies, diabetes and parkinsonism (Figure 2; www.mitomap.org May 2011; Ruiz-Pesini *et al.*, 2007). It has been estimated that the prevalence of mitochondrial disorders in the human population is 1/5000, which is considered to be relatively common for a metabolic disease (Schaefer *et al.*, 2004).

Although the mitochondrial genome encodes few proteins (13 in total, Figure 2), the mitochondrion is a protein-rich organelle, due to the import of nuclear-encoded proteins from the cellular cytoplasm (DiMauro & Schon, 2003). Thus, there is cross talk between the two genomes, meaning that mitochondrial disorders may have a nuclear or mitochondrial genetic origin.
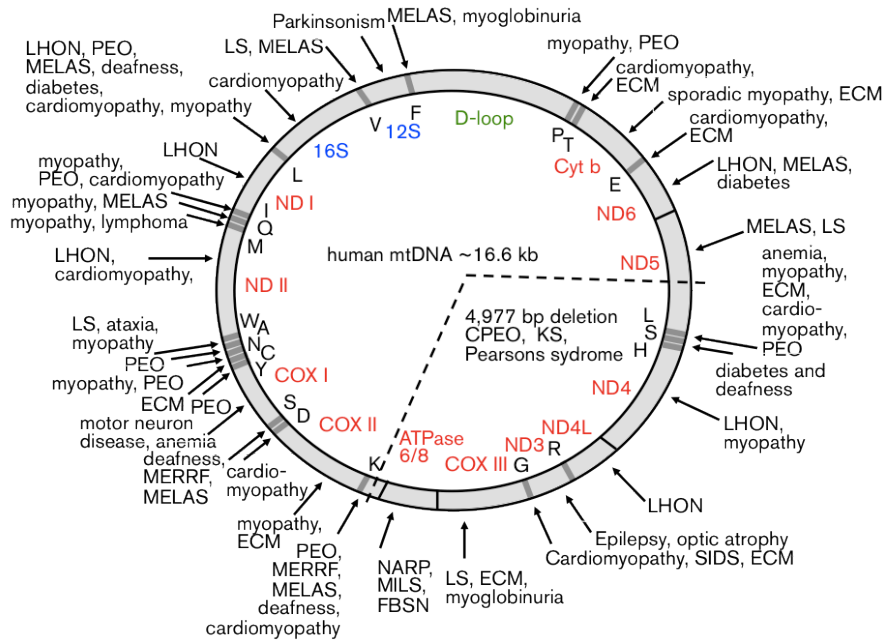
Figure 2. Schematic overview of the human mitochondrial genome including the approximate positions of several disease-causing mutations and their associated disorders. Red, protein coding genes; rRNA, blue; green, displacement-loop; black, *tRNA* genes and associated disorders. ECM, encephalomyopathy; FBSN, familial bilateral striatial necrosis; LOHN, Leber hereditary optic neuropathy; LS, Leigh's syndrome; MELAS, mitochondrial encephalomyopathy, lactic acidosis and stroke like episodes; MERRF, myoclonic epilepsy with ragged-red fibres; MILS, maternally inherited Leigh's syndrome; NARP, neuropathy, ataxia and retinitis pigmentosa; PEO, progressive external ophthalmoplegia; PPK, palmoplantar keratoderma; SIDS, sudden infant death syndrome (www.mitomap.org; Ruiz-Pesini *et al.*, 2007).

18

Three major aspects by which mitochondrial and nuclear genetics differ

There are three main differences between nuclear and mitochondrial genetics (DiMauro & Schon, 2001; DiMauro & Schon, 2003; DiMauro & Davidzon, 2005; Taylor & Turnbull, 2005; Greaves & Taylor, 2006; Park & Larsson, 2011). These are:

- *Heteroplasmy and threshold effect.* In contrast to nuclear encoded genes, which comprise one maternal and one paternal allele, there are hundreds to several thousand mtDNA molecules in each cell. The presence of one population of identical mtDNA molecules is called homoplasmy, whereas the existence of two or more populations of mtDNA is referred to as heteroplasmy. Heteroplasmic mutations are harmless unless the mutant load (proportion between wild type and mutant mtDNA) exceeds a certain threshold, which varies for different mitochondrial mutations and tissues.

- *Maternal inheritance.* Unlike the nDNA, the mtDNA is inherited strictly from the mother and transmitted along the maternal lineage, thus mtDNA lineages are clonal. Paternal mtDNA are degraded through an ubiquitin-mediated process (Sutovsky *et al.*, 1999; Sutovsky *et al.*, 2000). During meiosis only a limited number of mtDNA molecules are transmitted to the progeny due to a "bottleneck" effect (Jenuth *et al.*, 1996; Cao *et al.*, 2007). Consequently, the mutant load may shift rapidly between generations.

- *Random mitotic segregation* (Figure 3). Replication of mtDNA is not dependent on the cell cycle. Consequently, mtDNA molecules can replicate many times, or not at all, during a cell cycle. The degree of heteroplasmy may shift rapidly between different cell types and tissues. This may alter the clinical phenotype in an individual with time.
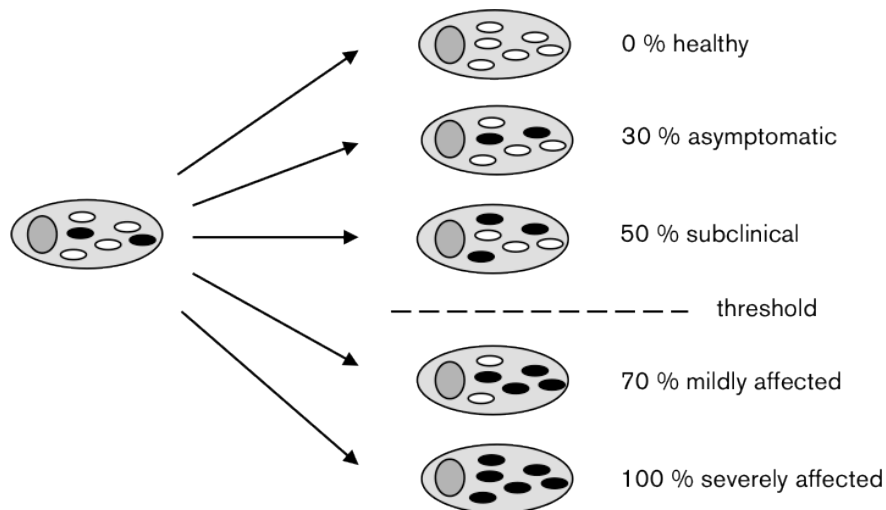
Figure 3. Random mitotic segregation of mtDNA. Upon mitosis mtDNA is randomly transmitted to daughter cells, hence each cell displays a different mutant load and clinical phenotype. The threshold defines the tolerated mutant load before deleterious clinical symptoms develop. Within a cell, light grey circle, white dots represent wild type mtDNA, black dots symbolize mutant mtDNA and the dark grey dot represents the nucleus.

Identifying the genetic origin of a mitochondrial disorder

Different approaches can be utilized to assess the genetic origin of a mitochondrial disorder; (I) Relatedness of cases can be traced in a pedigree. If all cases are traced back on the maternal lineage to the same female ancestor, there is a high probability that the disorder has a mitochondrial origin. However, due to the extensive generation time in humans, information from pedigrees comprised of several generations is limited. (II) Enzyme complex activity and ATP production, or oxygen consumption, of the respiratory chain can be measured. Since different complexes are encoded by nDNA and mtDNA, it may be possible to elucidate the origin of the disease and subsequently the mutation (Wibom *et al.*, 2002). (III) In a situation when there is no pedigree information available and biochemical measurements of the respiratory chain are ambiguous, perhaps the most informative strategy is to re-sequence the entire mitochondrial genome, which is relatively effortless due to its small size.

Assessing mitochondrial mutation causality

Several criteria have been established supporting the deleterious role of a novel mtDNA mutation (Naviaux, 2000; DiMauro & Schon, 2001; DiMauro & Davidzon, 2005). (I) The mutation should change an evolutionary and functionally important site. (II) Cases should be heteroplasmic for the mutation, although there are exceptions (Taylor *et al.*, 2003; Taylor & Turnbull, 2005). (III) The severity of symptoms in a family should correlate with the degree of heteroplasmy. (IV) Deficiencies of the respiratory chain enzymes or mitochondrial protein synthesis should be functionally confirmed (V) Histochemical staining should confirm pathogenicity.

Although many mitochondrial diseases have been correlated with mutations of the mtDNA, the relationship between phenotype and genotype is only partly understood (Graff *et al.*, 2002; Taylor & Turnbull, 2005). This could be due to a variety of factors including the different degrees of heteroplasmy in different parts of the organism possibly fluctuating with time; diverse nuclear genetic background; interaction with other mitochondrial genes, as well as environmental factors.

Apart from sensory ataxic neuropathy, reported in this thesis, only one additional spontaneously occurring mitochondrial disorder with comprehensive genetics has been reported in dogs, canine spongiform leukoencephalomyelopathy (Li *et al.*, 2006). However, many different mouse models for various mtDNA mutations exist (Wallace, 2001; Park & Larsson, 2011).

The dog as a model in genetic trait mapping

Population history of the domestic dog

The domestication of the dog (*Canis familiaris*) from the grey wolf (*Canis lupus*) traces back at least 15,000 years, but possibly as far back as 31,000 years (7,000-15,000 generations), and has been referred to as the first genetic bottleneck in dogs (Figure 4) (Lindblad-Toh *et al.*, 2005; Vonholdt *et al.*, 2010). At this point dogs displayed a large range of genetic diversity and breeding was not strongly influenced by man. Dogs and humans had a mutually beneficial relationship. Gradually, humans began to select dogs for certain traits such as desired behaviours (hunting, herding, retrieving) and appealing morphology (coat structure and colour, body size). Approximately 200 years ago (corresponding to 50-100 generations) one of the most extensive breeding experiments was initiated: breed creation, in which strict

breed standards were developed and applied. This was the second genetic bottleneck (Figure 4), selecting few founder animals of pre-breed dogs for creating each closed isolated population (present dog breeds). As a result we can observe reduced variation (phenotypic and genetic) within present breeds compared to that seen across breeds. As a population however, dogs show more morphological diversity than any other domesticated animal, with more than 400 breeds registered.

1st bottleneck: domestication
15,000-31,000 years ago
(7,000-15,000 generations ago)

2nd bottleneck: breed creation
approx. 200 years ago
(50-100 generations ago)

Wolves   Purebred dog isolates
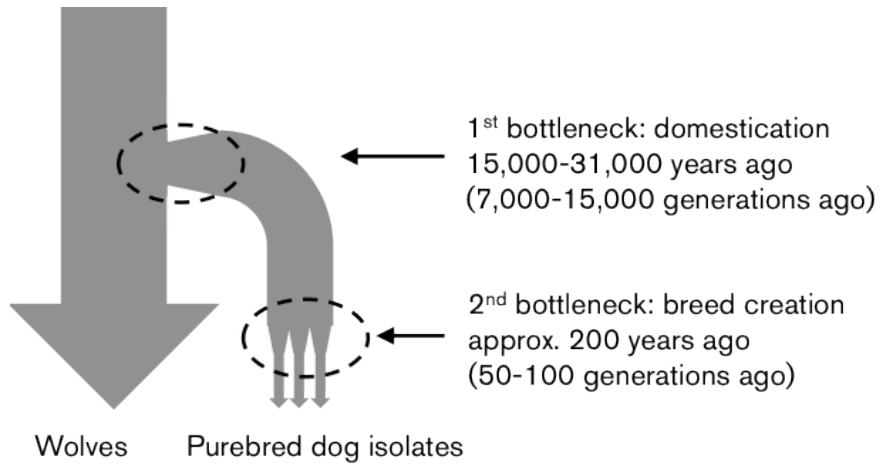
Figure 4. Population bottlenecks in the dog. The first population bottleneck occurred 15,000–31,000 years ago at the divergence from the wolf. At breed creation, the second genetic bottleneck (approximately 200 years ago) gave rise to pure bred dog isolates, each breed originating from few founder dogs (figure modified from Lindblad–Toh *et al.* (2005)).

Exploring the potential of the dog in whole-genome mapping studies

As a consequence of strict selective breeding and genetic drift due to the limited effective population size, unfavourable traits have become enriched in different breeds. These include diseases like cancers, neurological diseases (such as epilepsy), heart diseases, autoimmune and metabolic disorders (including diabetes), all of which are also common in humans (Ostrander *et al.*, 2000; Patterson, 2000; Sutter & Ostrander, 2004; Lindblad-Toh *et al.*, 2005; Peltonen *et al.*, 2006; Karlsson & Lindblad-Toh, 2008). The selective enrichment of risk alleles may be due to random fixation during the bottleneck events, hitchhiking (disease mutations located in close vicinity to alleles under positive selection) and also pleiotropy (a single mutation gives rise to several phenotypes). The high prevalence of certain diseases in a

limited number of breeds indicates that the causative mutation(s) occurred in the ancestral pre-breed dog population. Due to the limited number of founder animals used during breed creation it is likely that a limited number of disease-causing alleles have subsequently become enriched in the different breeds. Hence, detecting loci explaining the causality of a disease may be less challenging in dogs than in humans.

Apart from humans, dogs have the most extensive medical care as well as large records of documented relatedness. During the last decade disease-causing mutations of approximately 25 Mendelian disorders, including narcolepsy, kidney cancer and copper toxicosis, have been identified (Mignot et al., 1991; Lin et al., 1999; Jonasdottir et al., 2000; van De Sluis et al., 2002; Lingaas et al., 2003; Karlsson & Lindblad-Toh, 2008) utilizing tools such as linkage, radiation hybrid and comparative maps (Mellersh et al., 1997; Sutter & Ostrander, 2004). Developed as part of the Dog Genome Project, the relatively recent characterization of the high quality dog genome sequence, 7.5X coverage, offers improved possibilities for trait mapping and functional characterization in dogs (Lindblad-Toh et al., 2005). In addition to publishing the dog genome sequence, a comprehensive SNP catalogue was made available to the public, comprising of more than 2.5 million SNPs scattered along the genome (approximately 1 SNP/kb). Moreover, LD and haplotype blocks were defined within and between breeds. Consequently, within a breed haplotypes are large (0.5-1 Mb) and LD is often extensive (several megabases), about 10-100 times more extensive in dogs than in humans (Sutter et al., 2004; Lindblad-Toh et al., 2005). Between breeds, haplotypes are shorter and LD extends only tens of kb. This unique LD pattern is consistent with the occurrence of two distinct population bottlenecks (Lindblad-Toh et al., 2005; Karlsson & Lindblad-Toh, 2008).

The LD pattern and haplotype structure imply that fewer individuals and markers are needed in dogs, compared to humans, to perform GWAS. The number of dogs required in GWAS varies, depending on the inheritance pattern (Lindblad-Toh et al., 2005; Karlsson & Lindblad-Toh, 2008). It has been estimated that for mapping a recessive trait with high penetrance 20 cases and 20 controls is sufficient, whereas for a dominant trait, 50 cases and 50 controls should provide sufficient power. Approximately 100 cases and 100 controls are needed to detect alleles with 5-fold increased risk. It has been estimated that 10,000-15,000 SNPs (Lindblad-Toh et al., 2005) should be sufficient for most purposes, compared to 300,000 to 1,000,000 SNPs in humans (Gabriel et al., 2002; Consortium, 2005). Moreover, it has been assessed that within a breed, on average 73 % of the SNPs are informative

(polymorphic) with a minor allele frequency (MAF) 5-50 % (Lindblad–Toh *et al.*, 2005).

Hence, based on the structure of the dog genome, GWAS should be powerful for mapping both monogenic and complex traits with fewer SNPs and a smaller sample size compared to humans. Moreover, the strikingly similar clinical phenotype between dogs and humans suggest that the dog may become man's best friend in providing new insights into the biology of diseases shared by dogs and people.

Dogs are not only a potential model organism for mapping traits with nuclear origin; they also offer a model for studying spontaneously occurring mitochondrial disorders. While the first mitochondrial disease-causing mutations were identified in humans more than two decades ago, the advances in dogs have been limited. Several clinical characterizations suggesting mitochondrial origin have been published, however, only two have identified the causative mutation (Li *et al.*, 2006; Paper II). Hence, the potential of utilizing the dog as a model for mitochondrial disorders is still in its infancy.

Aims of this Thesis

The main objective of this thesis was to assess the potential of the dog as a model for mapping traits with nuclear and mitochondrial origin.

The specific aims were to:

(I) Map the white spotting locus in dogs by utilizing two-stage association mapping and also to provide functional evidence for the identified coat colour candidate mutations. (Papers I and III)

(II) Identify and functionally characterize the mutation causing sensory ataxic neuropathy in Golden Retrievers. (Paper II)

Present Studies

MAPPING TWO TRAITS CONTROLLED BY TWO DIFFERENT GENOMES

White spotting locus - proof-of-principle study (Papers I and III)

Background

The dog genome paper (Lindblad-Toh *et al.*, 2005), including the high-quality draft genome sequence, a dense SNP map (comprising more than 2.5 million SNPs) as well as the structure of genetic variation in dogs, led to the next inevitable step of assessing the full potential of the dog as a model organism for genetic trait mapping. The extensive LD and large haplotype blocks within breeds suggests that few individuals and a relatively sparse marker set can be used to scan the genome for association in one breed. Based on the simulated power predictions by Lindblad-Toh *et al.* (2005) it was estimated that approximately 20 affected and 20 controls should be sufficient to map a recessive trait. The short LD and short haplotype blocks across breeds indicate that it should be feasible to decrease the region of association identified in one breed by adding additional breeds presenting the same phenotype and sharing the same ancestral origin for the mapped trait. Thus, a two-stage approach should be efficient to map trait genes in dogs.

In the first paper of this thesis we describe the development and general characteristics of the first generation canine SNP array. Moreover, we analysed the previously estimated LD and haplotype blocks on genome-wide level and assessed breed relationships. To prove the efficiency of using the dog as a model we mapped two Mendelian traits: the extreme white coat colour in Boxer and the ridge in Rhodesian Ridgeback (Paper I). In

the third paper of this thesis, the identified candidate mutations of the white spotting locus were functionally evaluated and further assessed.

Results and Discussion

Brief summary of canine SNP arrays and assessment of LD and haplotype blocks

In order to be able to perform GWAS in dogs we developed the first generation canine Affymetrix array containing 26,578 SNPs (27 K), and evaluated its performance in GWAS. Three additional SNP arrays have been developed since this paper was published, the 50 K (49,663 SNPs) Affymetrix array and two Illumina arrays, a 22 K (22,362 SNPs) and a 170 K (174,943 markers) SNP chip. SNPs included on the four arrays have been chosen to maximize overlap between arrays potentially allowing meta-analyses across different arrays and platforms. There are different advantages to the various designs, although all are functional for GWAS (Karlsson & Lindblad-Toh, 2008; Vaysse *et al.*, 2011). Due to the 170 K Illumina arrays uniform genome coverage, high call rate and much higher SNP density it may be preferable as it, in addition, enables GWAS across breeds, as well as allows for mapping of selective sweeps and CNVs.

The theory of breed development (Figure 4) as previously reported suggests that different haplotype structure should be found between and within breeds. We confirmed this on a genome-wide level by analysing the haplotype structure in 250 dogs using the 27 K array. We found that (I) within breeds haplotype blocks are long and LD is extensive, as compared to short haplotypes and LD dropping off rapidly with distance between breeds. (II) The genetic differentiation between breeds is high, which reflects the tight bottlenecks at breed creation. (III) We noticed that there is considerable genetic differentiation between geographically separated populations of the same breed, and suggest that this could be explained by genetic drift.

White spotting locus successfully mapped with a two-stage approach

The white spotting locus (*S*) in dogs was defined by Little (1957). He described four different alleles: solid or wild type (*S*, single coloured coat without spots), Irish spotting (*S^i*, moderate white spotting), piebald (*S^p*, extensive white spotting) and extreme white (*S^w*, complete white coat) (Figure 5).
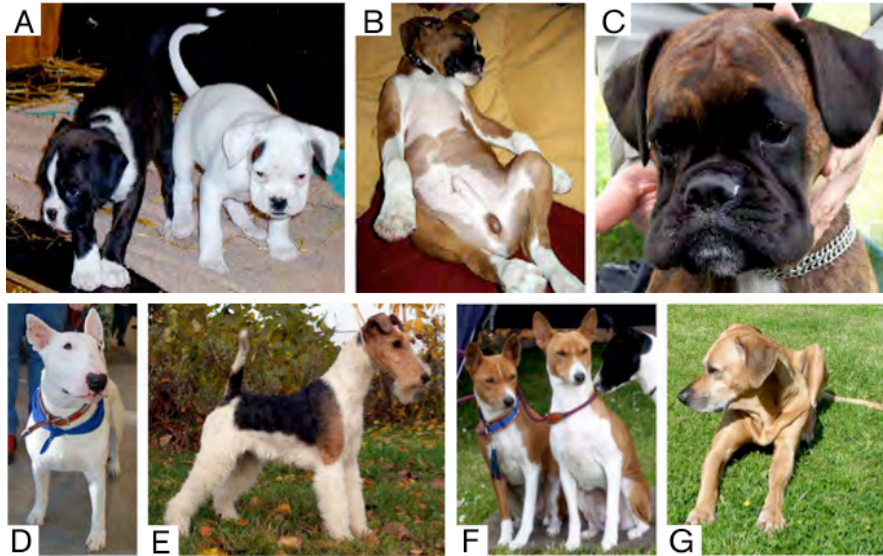
Figure 5. Coat colour phenotypes at the white spotting locus. (A) Left, flash Boxer (S^w/S) and right, white Boxer (S^w/S^w), (B) flash phenotype in Boxer (S^w/S), (C) solid Boxer (S/S). Coat colour phenotypes in other breeds. (D) White Bull Terrier (S^w/S^w), (E) piebald Fox Terrier (S^p/S^p), (F) Irish spotting in Basenji (S^i/S^i), (G) solid Rhodesian Ridgeback (S/S) (figure modified from Paper I).

In the Boxer breed two of the alleles exist: S and S^w, giving rise to three different phenotypes (Figure 5 a-c): solid (S/S), flash or semi–Irish (S^w/S) and extreme white (S^w/S^w). This semi–dominantly inherited trait will be further discussed to demonstrate our two–stage strategy for trait mapping in dogs.

In the first step of our strategy we utilized ten white (S^w/S^w) and nine solid Boxers (S/S), taking advantage of the long haplotypes and extensive LD within breeds (Figure 6). With relatively few SNPs (27 K array), a genome-wide significant peak of less than 1 Mb was identified on chromosome 20. The most associated SNP was within an approximately 800 kb haplotype block, defined by 11 SNPs. The association for this region was 1000-fold stronger than for any other region of the genome and the region contained only one gene, *microphtalmia-associated transcription factor* (*MITF*). The *MITF* gene proved to be an excellent candidate since MITF is a crucial transcription factor affecting melanocyte migration and development. *MITF*

loss-of-function mutations in humans and mice usually have deleterious pleiotropic effects, affecting for example pigmentation, hearing and vision (Tassabehji *et al.*, 1994; Smith *et al.*, 2000; Steingrimsson *et al.*, 2004). In dogs it seems to have a limited effect on other phenotypic traits, since only about 2 % of white dogs (S^w/S^w) present with bilateral deafness. Furthermore, the fact that other pigmented areas of white spotted dogs appear to have normal pigmentation suggests normal melanocyte function in melanocytes that reach the dermis. Hence, regulatory mutations of *MITF* are more plausible than coding mutations, as functional MITF is crucial for pigment production.

In order to decrease the associated region in Boxers, we performed finemapping (second stage mapping), by adding a second breed displaying the same phenotype, Bull Terriers (Figure 5D). In total 115 SNPs, spanning a region of 4.6 Mb, were genotyped in 127 Boxers and Bull Terriers (*S/S*, S^w/S and S^w/S^w). Sixty-eight SNPs were in the associated 800 kb haplotype and the remainder in flanking regions. First the association was performed in each breed separately, then the two association peaks were compared and the minimum shared haplotype was identified. Haplotype analysis revealed a 102 kb region, which included two haplotype blocks with perfect genotype-phenotype correlation in both breeds. One of the blocks (approximately 20 kb) included the melanocyte-specific 1M promoter, crucial for melanocyte function, and the other (about 10 kb) contained exon 1B, which is included in several different MITF isoforms. Interestingly, the associated blocks were separated by a middle region of approximately 40 kb, which indicates a site of recombination.
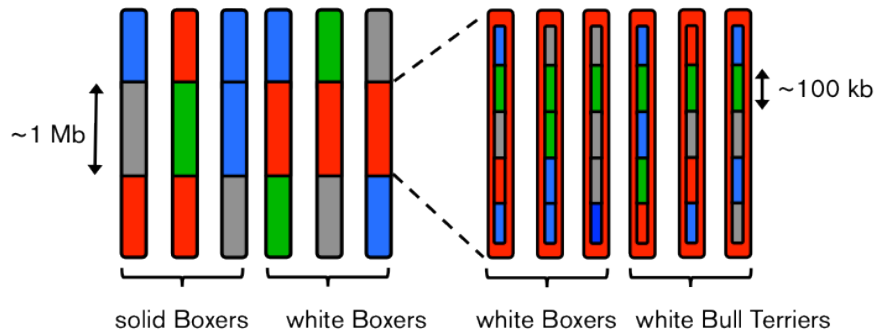
Figure 6. Two-stage strategy. In the first step a trait is mapped in one breed, taking advantage of the long haplotype and extensive LD. A region less than 1 Mb is identified. Additional breeds presenting with the same phenotype are used in the second step to reveal the ancient shared haplotypes. Thus the region is narrowed down to approximately 100 kb (figure modified from Karlsson & Lindblad-Toh (2008)).

Candidate mutations in MITF

To identify candidate mutations we re-sequenced two BAC clones, derived from the sequenced flash coloured Boxer used to produce the dog genome assembly. These two BAC clones represented the solid and white haplotypes respectively. In the 102 kb associated region we identified 124 sequence differences all located in non-coding regions, indicating that the extreme white phenotype is caused by regulatory mutation(s). Candidate mutations were evaluated in a larger material of white, flash and solid Boxers and Bull Terriers, as well as in additional solid dogs. Seventy-eight mutations were excluded since they were discordant with the coat colour phenotype. The remaining 46 were evaluated based on their conservation across species, leaving three top candidates. (I) The first candidate was a short interspersed nucleotide element (SINE) insertion (SINEC–Cf element) located approximately 3 kb upstream of the melanocyte specific 1M promoter. The SINE was present in all white and piebald dogs (homozygous), heterozygous in flash Boxers and Bull Terriers and absent in Irish and solid dogs. This was a plausible candidate since SINE elements have been reported to affect transcriptional regulation. (II) The second candidate was a length polymorphism (Lp) of a short set of nucleotide repeats located approximately 100 bp upstream of 1M promoter. White Boxer and Bull Terriers displayed alleles of 35 bp, which is 4 bp longer than the allele in

solid Boxers and Bull Terriers. The long variants of 35–36 bp occurred in white, piebald and Irish spotted breeds, whereas the short variant occurred in solid dogs. Interestingly, there was larger sequence variability among the long variant, with six alleles in six breeds, as compared to four alleles of the short Lp in 12 breeds. This possibly reflects positive selection for the variants affecting the pattern of white spotting. Moreover, the length of the Lp may affect the interaction of transcription factors upstream and downstream of the Lp. (III) The third candidate was a SNP located approximately 1.200 bp upstream of the 1M promoter of *MITF* indicated at position –1.100 bp at number 21 in Suppl. Table 3a in Paper I. Although it is located close to conserved elements, the SNP is not conserved across mammals, hence it is unlikely to affect this coat colour phenotype. Thus, we pursued functional studies to define the regulatory effects of the SINE insertion and the length polymorphism (Figure 7).
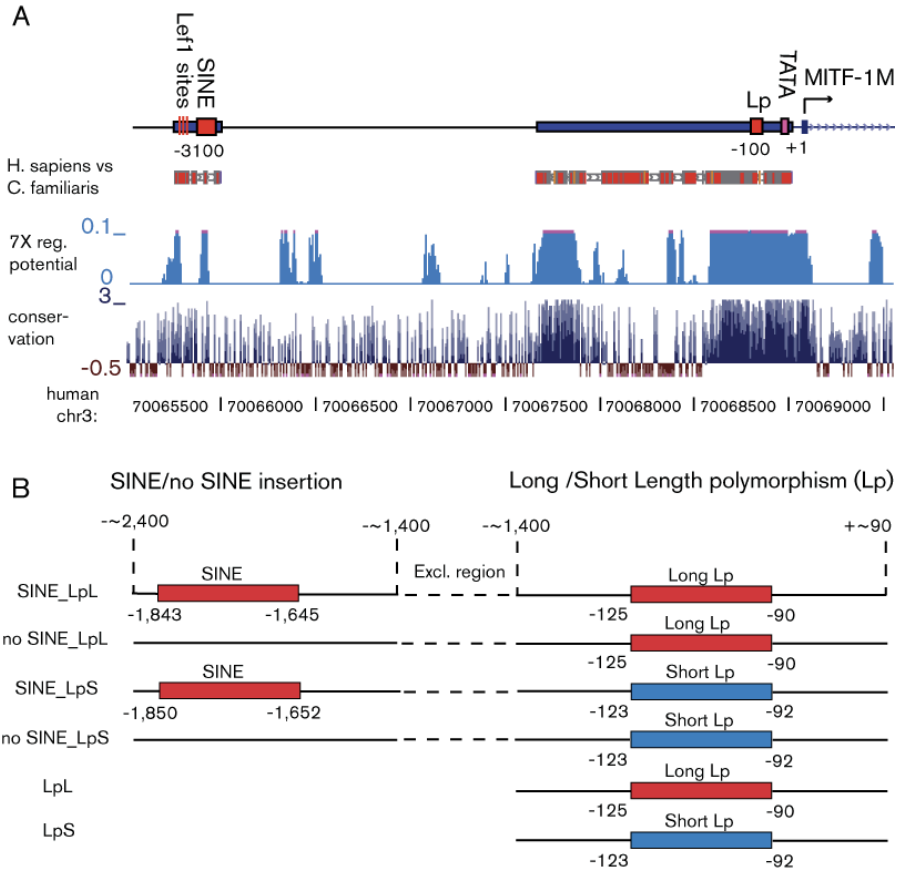
Figure 7. Two candidate mutations, a SINE insertion and a length polymorphism (Lp), identified upstream of the exon 1M promoter. (A) The regions defined including a comparative human–dog alignment over the region, the 7X regulatory potential and conservation across species (UCSC genome browser, Human Mar. 2006 assembly, NCBI36/hg18). (B) The six different reporter constructs included in the Luciferase assays. Numbering refers to the position in the different inserts.

Reporters suggest complex regulation of MITF

The functional consequences of the SINE insertion and the variable Lp were investigated utilizing a Dual Luciferase Reporter Assay System. Six different constructs were designed as illustrated by Figure 7: SINE + long Lp (LpL), corresponding to extreme white allele, no SINE + LpL, similar to

33

Irish spotted allele, SINE + short Lp (LpS) and, no SINE + LpS, corresponding to the solid allele. We also included two constructs with just the long or short Lp. The additional sequence included in the constructs, about 800-1000 bp upstream and downstream of the SINE, and about 1.5 kb upstream and downstream of the Lp, included potential regulatory elements in the region based on the estimated 7X regulatory potential and across species conservation presented in the UCSC Genome Browser (Kolbe *et al.*, 2004; King *et al.*, 2005)(http://genome.uscs.edu/). Hence other previously identified polymorphism were included in these flanking sequences, including the third top candidate SNP approximately 1.200 bp upstream of transcriptional start site, as shown in Table 1.

Table 1. Polymorphisms in Luciferase insert design including SINE and Lp. Additional polymorphisms that are included in the design, that we considered unlikely to be functionally important, are also presented, including relative position in relation to cap site, phenotype correlation, reason for exclusion and numbering according to Suppl. Table 3a (Paper I).

Sequence polymorphism	Position relative to cap site, +1	Phenotype correlation	Reason for exclusion	# in Suppl. Table 3a Paper I
Simple repeat (12/14 bases)	-~3,450 bp	No	a	25
SINE insertion in white	-~3,150 bp	Yes		24
Candidate SNP	-~1,200 bp	Yes		21
Simple repeat (9 bases)	-~700 bp	No	a	20
SNP	-~600 bp	No	a	19
Length polymorphism	-~100 bp	Yes		18
Indel (2 bases)	-~60 bp	No	a	17
SNP (novel)	+ 86 bp (UTR)	Unknown		

Total size of construct (white): ~2.500 bp

a, phenotype discordance

SINE sequence insertion (white): ~-3.800 - ~-2.800 bp (~1.000 bp)

Excluded region from design: ~-2.800 - ~-1.400 bp (~1.400 bp)

Lp sequence (white): ~-1.400 - ~+90 bp (~1.500 bp)

Luciferase activity in the mouse melanocyte melan-a cell line and human melanoma cell line 624mel showed an approximately 100-fold difference in expression between the two cell lines. Due to suboptimal transfection efficiency, results from melan-a cell line were excluded. Thus, we focused on interpreting the results from 624mel cell line showing higher and more reliable transcriptional activity in the reporters tested.

Reporters with the long Lp showed a considerable decrease in promoter activity compared to reporters with the short Lp in the 624mel cell line (Figure 8). Thus, the long variant of the Lp consistently repressed the reporter activity in 624mel cell line. A lower activity was observed in constructs with the SINE insertion as compared to reporters lacking the SINE (Figure 8, SINE_LpL vs no SINE_LpL and SINE_LpS vs no SINE_LpS). There was a 3-fold difference between the constructs representing white (SINE_LpL) and solid (no SINE_LpS) haplotypes. The construct similar to the Irish spotting haplotype (no SINE_LpL) showed intermediate reporter activity (Figure 8).
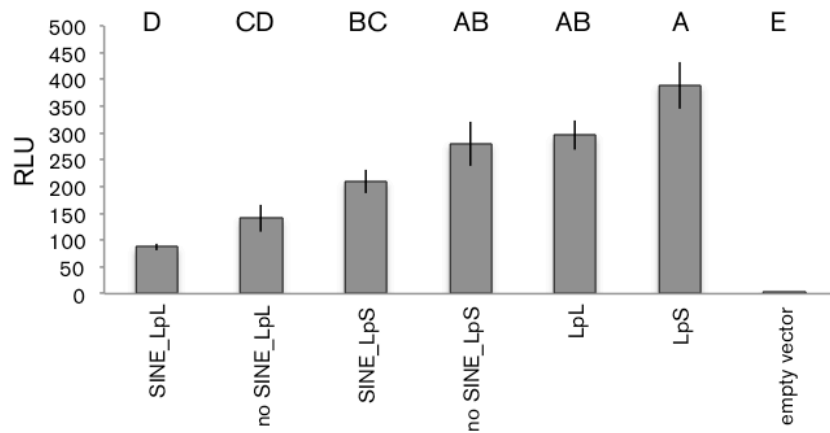
Figure 8. Reporter activity in human melanoma 624mel cell line. Firefly luciferase levels normalized against control vector Renilla luciferase. Measurements were normalized against the empty vector, and each experiment was repeated three times, each sample was measured in triplicate. Pair-wise comparisons of constructs with long length polymorphism (LpL) and short length polymorphism (LpS) show decreased expression in constructs with LpL. Pair-wise comparisons of constructs with SINE compared to those without, show decreased promoter activity in reporters where the SINE is included. Capital letters above each bar show the significance (p<0.001) in promoter activity in pairwise comparisons between the different reporter constructs tested. Reporters that do not share a letter show a significant difference. Error bars represent standard error of the mean. RLU=relative luciferase unit

Potential regulatory elements were analysed in regions in proximity to the SINE insertion and Lp using TRANSFAC Professional (Matys *et al.*, 2006). Approximately 250 bp upstream of the SINE insertion a cluster of Lef1 sites was identified. Their functionality was evaluated and confirmed by electrophoretic mobility shift assay (EMSA). However, there are additional predicted transcriptional regulatory elements in the SINE and Lp regions that need to be further assessed.

Concluding remarks of white spotting locus and MITF

Data presented in these studies prove the efficiency of using a two-stage approach for mapping monogenic traits in dogs. In the first stage, a 1 Mb region was identified in one breed taking advantage of the large breed-

specific haplotypes. In the second stage, an additional breed was included revealing a shared haplotype block, thus the region was decreased to approximately 100 kb. Mutation screening including additional breeds revealed at least two plausible candidate mutations that were assessed further. Moreover, the current study successfully utilized the 27 K array, suggesting the SNP density should be sufficient in other dog genetic studies.

The data presented in these studies suggest that S^w, S^p and S^i do not represent three independent mutations at the same locus, but rather different combinations of a set of mutations at the same locus because of the haplotype sharing we observe among the white spotting alleles. We believe that the combination of different regulatory mutations will define the extent of pigmentation. This differs from mice where most *Mitf* alleles represent a single mutation with deleterious effect, such as depigmentation, hearing loss and various eye and bone diseases. Phenotypically, the black-eyed white *Mitf mi-bw* mouse model resembles the extreme white phenotype in dogs. This mouse phenotype is caused by an insertion in intron 3 which affects *MITF* expression and melanocyte development, causing individuals to be both white and deaf. The extreme white allele in dogs gives rise to a mildly deleterious phenotype, with approximately 2 % of the dogs presenting with bilateral deafness (Strain, 2004). This indicated that *MITF-M* expression in dogs is sufficient for limited melanocyte migration. The few coloured patches on white dogs show normal pigmentation, indicating that fully functional MITF-M is expressed in mature melanocytes. Our expression studies in human melanoma 624mel cell line revealed that different combinations of the SINE and Lp affect gene expression. However, further studies are necessary to unravel the complete functional consequences of these mutations.

Future prospects

Future studies of MITF

The most challenging aspect in the process of unravelling the genetic cause of white spotting in dogs, lies in the identification and functional confirmation of the causative mutations in the *MITF* gene. MITF belongs to the Myc supergene family of basic helix–loop–helix (bHLH)–leucine zipper transcription factors and it is highly conserved in vertebrate species (Lin & Fisher, 2007). In humans, nine different isoforms have been identified, mainly differing in their 5' region; exons 2 to 9 are common in all nine isoforms (Levy *et al.*, 2006). These isoforms are expressed in different tissues, MITF-M shows melanocyte-specific expression. The SINE

and Lp identified upstream of the MITF-M promoter are indeed probable candidates but it may be possible that there are other yet unidentified polymorphisms that also contribute to the white spotting phenotype.

Reporters utilized in the Luciferase assays strongly support the causality of the SINE and Lp sequences. However, this experiment could be complemented by additional constructs. It would be interesting to design reporters which include the entire 4 kb region upstream of promoter 1M, mimicking the "natural situation" in the white, piebald, Irish and solid haplotypes and to use them in transient transfection experiments to functionally evaluate the entire region. If a difference in expression would be observed between the S^w and S^p haplotypes, it would reveal the presence of additional unidentified regulatory elements, explaining the difference between S^w and S^p phenotypes which cannot be explained by the SINE and Lp alone. Moreover, reporters including only the SINE and Lp would show if they specifically function as transcription factor binding sites. Also, creating a dog melanocyte cell line may be considered as it could potentially prove to be a more optimal system for further evaluating the functional consequences of the causative white spotting polymorphisms.

MITF is an important developmental gene expressed during early embryogenesis. This complicates functional studies in tissue samples from dogs as one would need to obtain dog embryos, which is not a trivial task. Since it has been shown that SINEs may be targeted for methylation (Arnaud *et al.*, 2000), it would be interesting to investigate if the degree of methylation of the SINE insertion in S^w and S^p dogs differ. Thus, this could potentially explain the difference between the two phenotypes, which both contain the SINE and the long variant of the Lp.

Sequence analysis of SINE-region revealed a cluster of Lef1 sites which were shown to form complexes with nuclear extracts from human melanoma 624mel cell line and mouse melanoma B16 cell line. Additional regulatory elements were predicted in both the SINE and Lp regions which need to be further investigated and functionally assessed.

A third candidate polymorphism, a SNP (nr 21 Suppl. Table 3a, Paper I), was considered less likely to affect the white spotting phenotype due to the low degree of conservation across species. A possible way to assess its functional importance is to mutate that position in the available constructs that have already been tested and perform Luciferase assays and compare the obtained activities. If the expression in the original and mutated constructs is similar one can exclude the causality of this SNP in white spotting.

Finally, the dog model offers a complement to the already existing mouse models to further unravel the molecular mechanisms and cellular pathways

in pigmentation biology. It may also serve as a valuable model for MITF-related disorders in humans since for example deafness (bilateral and unilateral) is linked to the extreme white phenotype in dogs.

Challenges and potential solutions in genetic trait mapping in dogs

At the time that our two-stage strategy was published, it brought great hopes and expectations, not only to the dog community but also to disease geneticists as they expected that the dog model could offer a shortcut and shed light on the genetics behind disorders in humans. Although the second expectation has been met, along the way trait mapping in dogs has proved to be more challenging than first anticipated.

GWAS are most powerful when unrelated cases and controls are used. The optimal sample size depends on the predicted mode of inheritance and how strong effects the risk factors have. Insufficient numbers of cases and controls will decrease the probability of identifying a true region of association. Due to the complicated family relationships in purebred dogs, finding and sampling a material that meet these demands can be challenging. Moreover, correct clinical characterization of both cases and controls is the first crucial step that cannot be over looked. Well-defined clinical inclusion and exclusion criteria for cases and controls are required to obtain this goal. In some projects it has been more challenging to collect healthy controls than cases, since for some diseases a large proportion of the population belong to the "grey-zone". A possible way to increase the number of samples is to include available cases and controls from different countries. In fact, a few years ago the LUPA consortium was initiated with the specific aim to unravel common human diseases using dog genetics. Partners of the consortium exchange samples and resources potentially allowing more efficient progress in the different studies.

An additional problem in GWAS lies with population stratification. As we showed in Paper I, different populations from the same breed may show genetic differentiation. Therefore, it is important to sample geographically matched cases and controls. Thus, the same number of cases and controls with similar relatedness should be collected from different countries when used in the same study. Due to the difficulty to meet these aspects, population stratification is thoroughly investigated and the development of tools to correct for it is highly prioritized as it could allow us to include more dogs.

The two-stage strategy that we presented in the first paper of this thesis is very powerful for mapping loci that have been under strong selection, such as coat colour, where it is expected that mutations or combinations of old

risk factors are identical by descent, i.e. that they were fixed prior to breed creation. However in several studies of complex diseases, we have observed that although different breeds present similar clinical disease phenotype, they either: (I) could have different underlying risk loci, which may suggest that the risk loci arose after breed formation and possibly after the divergence of one single breed into different subpopulations. (II) Alternatively, the disease-causing mutations arose in the pre-breed population and were present at low frequency in the ancestral population. During breed formation different mutations were enriched in different breeds due to the use of few founder dogs for each breed. The reason why we find association in different regions in different breeds presenting the same disease phenotype could be because the allele frequencies in different breeds differs or that the disease allele was only transmitted to a single breed in the first place.

It is also possible that one breed can have a sweep within another breed's risk-factor peak. Due to the extensive regions of homozygosity, the association may not be detected. Once a causative mutation has been identified and confirmed in one breed, it would be interesting to investigate if the same mutation is present in breeds presenting the same phenotype but which show no association for that specific locus. This would reveal if the mutation arose before or after breed creation. Also, the risk factors between breeds may be shared, but the proportion of how much of a phenotypic trait in different breeds they explain may vary.

Despite these challenging aspects, several additional successful studies using the dog as a model to identify genetic risk factors for disease have been published. For example, the successful identification of risk loci underlying a canine SLE-related complex disorder in Nova Scotia Duck Tolling Retrievers (Wilbe *et al.*, 2010), a degenerative myelopathy similar to human ALS (Awano *et al.*, 2009) and Shar-Pei fever with a similar phenotype in humans (Olsson *et al.*, 2011). Conclusively, the greatest lesson we can learn from these studies is that each study offers its own challenges, and adjustments have to be made accordingly, but a lot of new biology is nonetheless learnt in an effective manner.

Sensory ataxic neuropathy - a mitochondrial disorder in dogs (Paper II)

Background

While more than 550 disease-associated polymorphisms of the mitochondrial genome have been reported in humans (www.mitomap.org; Ruiz-Pesini *et al.*, 2007) in dogs, the molecular genetic and biochemical data of suspected mitochondrial disorders is largely unknown. To date, there are only two mitochondrial disorders with mitochondrial origin in dogs for which disease-associated mutations have been identified and evaluated. The first one is canine spongiform leukoencephalomyelopathy (Li *et al.*, 2006) and the second is sensory ataxic neuropathy (SAN) in Golden Retrievers (Paper II).

SAN is a recently identified neurological disorder in Golden Retrievers with an insidious onset between two and eleven months of age, followed by a slow progression (Jaderlund *et al.*, 2007; Jaderlund, 2009). Males and females are affected at similar frequency and to our knowledge, there are approximately 30 cases in total worldwide.

Sensory ataxic neuropathies in humans are characterized by loss of proprioceptive sensation, loss of tendon reflexes whilst muscle mass and strength is preserved (Illa *et al.*, 2001). Dogs presenting SAN show similar symptoms (Jaderlund *et al.*, 2007), they are ataxic, display postural reaction deficits and have absent or reduced spinal reflexes. There are no detectable indications of muscle atrophy or pain. Affected dogs do display a minor reduction in conduction of nerve impulses in sensory nerves. Post-mortem examination has revealed mild degenerative changes in both central and peripheral nervous systems. Thus, canine SAN may be a useful model for sensory ataxic neuropathies affecting humans.

Results and Discussion

Pedigree analysis and candidate mutation suggest mitochondrial origin of SAN

In early 2002, Golden Retrievers presenting a similar neurodegenerative disease phenotype were brought to the University Clinic, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden by their owners. During the next few years, 25 affected dogs were diagnosed and subsequently traced on the maternal lineage to the same female ancestor that lived during the 1970s. This suggested a mitochondrial mode of disease inheritance. The proportion of affected progeny derived from female or male parents representing this maternal lineage was assessed. If all cases were

derived from female dogs, a mitochondrial DNA (mtDNA) risk factor was highly probable. In total there were 272 progeny from female dogs and 177 progeny from male dogs belonging to this maternal lineage, all 25 cases were among the 272 offspring derived from the females. Thus, we hypothesized that SAN is a neurodegenerative disorder with a mitochondrial origin.

By re-sequencing the complete mitochondrial genome (approximately 16.6 kb) of affected dogs, close relatives (defined as first, second, third or fourth degree relatives), and unrelated, unaffected Golden Retrievers, a one base pair deletion (ΔT5304) was identified in the mitochondrial $tRNA^{Tyrosine}$ ($tRNA^{Tyr}$) gene of affected dogs and their close relatives. This specific deletion has not been found in any other dog breed and the deleted position is highly conserved among vertebrates. While several disease-causing mitochondrial $tRNA^{Tyr}$ mutations have been reported in humans, none represents this specific deletion (Pulkes et al., 2000; Raffelsberger et al., 2001; Sahashi et al., 2001; Scaglia et al., 2003).

Evaluation of methods for quantification

Since clinically healthy relatives tested positive for this deletion, heteroplasmy was strongly suspected. Due to our inability to differentiate between the mutant load observed among affected dogs and healthy relatives, we evaluated the accuracy and resolution of two different methods: (I) pyrosequencing and (II) quantitative oligonucleotide ligation assay (qOLA). Pyrosequencing had been utilized previously for quantification purposes of mitochondrial heteroplasmies (Ronaghi & Elahi, 2002). Whereas, qOLA was recently presented by Seo et al. (2007) with the specific aim of analysing copy number variations (CNVs) more reliably and accurately than previously reported. Thus, this method had not been used to define heteroplasmy in mtDNA until now.

By cloning amplified PCR fragments from homoplasmic wild type and affected dogs, we created our own heteroplasmic dilution curves from homoplasmic clones. Two dilution series, defined as wide (0 % wt, 10 % wt, 20 % wt, 30 % wt... 100 % wt) and narrow (0.5 % wt, 1 % wt, 2 % wt, 3 % wt... 10 % wt) allowed evaluation of the two different methods (Figure 9). Triplicates within one experiment showed less variation in qOLA compared to pyrosequencing. Overall, qOLA showed higher accuracy than pyrosequencing and we estimated that we should be able to detect samples with as little as 2-3 % of wt mtDNA.
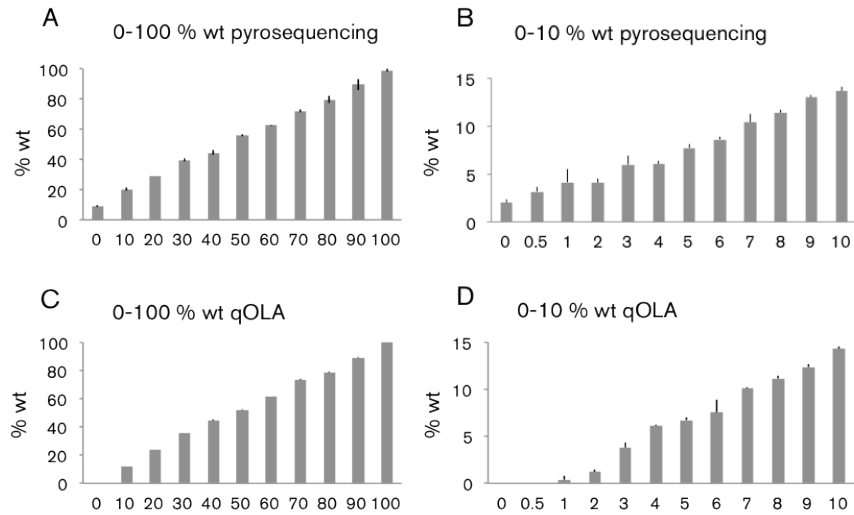
Figure 9. Dilution curves evaluated by pyrosequencing and quantitative oligonucleotide ligation assay (qOLA). (A) and (B) Pyroseqencing. (C) and (D) qOLA. Wide dilution series analysed with qOLA (C) shows less variation with a sample of triplicates as well as more accurate quantification of extreme values (0 % and 100 %) as compared to (A) pyrosequencing. Narrow dilution series of qOLA show higher accuracy compared to pyrosequencing. Error bars represent standard deviations of triplicates within one quantification (figure modified from Paper II).

Assessment of mutant load and functional evaluation of candidate deletion

Due to the higher accuracy and resolution, assessment of mutant load in blood samples was performed using qOLA (Figure 10). Samples from the following three groups of related Golden Retrievers were analysed: (I) affected dogs (n=20), (II) clinically healthy close relatives such as mothers, grandmothers and siblings of affected dogs (n=13), (III) distant relatives (n=8) descending from maternal lineages one and two generations above our suggested founder female. We also included healthy unrelated Golden Retrievers (n=71) and 86 dogs representing 18 breeds as well as six wolves. All unrelated dogs and wolves were homoplasmic for the wt allele, whereas the remaining three groups were clearly heteroplasmic. Distant relatives showed a large degree of variation in mutant load, ranging from 5-65 %, hence the female we assumed to be the founder was in fact not. The affected dogs showed very low levels of the wt allele, ranging from 0–11.2

%. Surprisingly some clinically healthy, close relatives showed comparable values. Our results suggest that the tRNATyr deletion is causative, as all cases reported descend from the branch with the highest mutant load. The imperfect correlation between disease phenotype and heteroplasmy in blood, could be explained by the fact that SAN is manifested in CNS (central nervous system) and not primarily in blood. Furthermore, the nuclear genetic background as well as environmental factors, could influence the penetrance of the disease (Graff et al., 2002; Taylor & Turnbull, 2005). Heteroplasmy was also assessed in tissue samples harvested post-mortem from three affected dogs and compared to the heteroplasmy measured in blood. All tissues showed higher mutant load compared to blood. Unfortunately, we did not have access to the corresponding tissues from related, clinically healthy dogs with similar mutant load in blood.
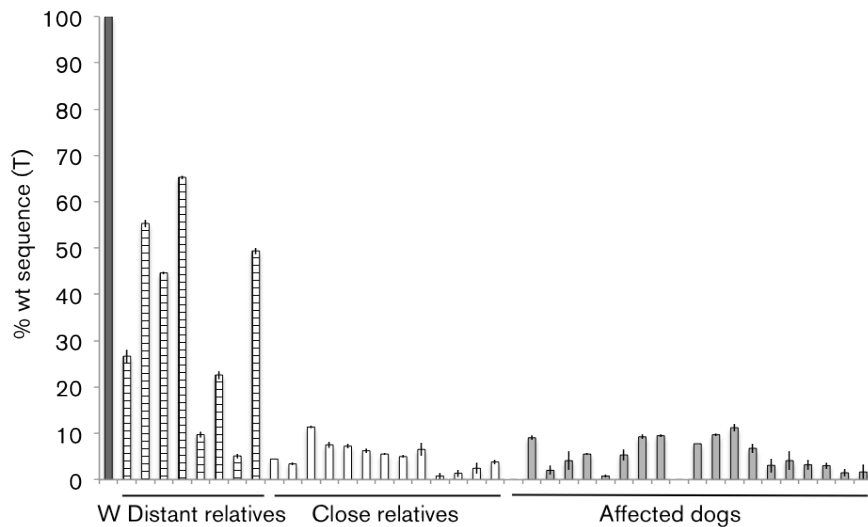
Figure 10. Evaluation of heteroplasmy in blood using quantitative oligonucleotide ligation assay. Dark grey bar (W) represent healthy unrelated Golden Retrievers (n=71). Degree of heteroplasmy is represented by striped bars in distant relatives, white bars represent close relatives and affected dogs are represented by light grey bars. There is an overlap in mutant load between affected Golden Retrievers and their close relatives. Error bars represent standard deviations of triplicates within one quantification (figure modified from Paper II).

44

Northern blots were used to assess the steady-state levels of $tRNA^{Tyr}$ from $tRNA$ extracted from the muscle tissue of two controls and three affected SAN dogs. Extracts from affected dogs showed 10-fold lower expression levels of $tRNA^{Tyr}$ as compared to controls. Equal expression of control $tRNAs$, $tRNA^{Cysteine}$ ($tRNA^{Cys}$) and $tRNA^{Glutamine}$ ($tRNA^{Gln}$), were detected in all five dogs. The lower expression of $tRNA^{Tyr}$ could possibly be explained by a processing defect during $tRNA^{Tyr}$ formation and maturation. However, this is not very likely since $tRNA^{Tyr}$ and $tRNA^{Cys}$ are transcribed as one polycistronic precursor, hence processing defects should affect both. Moreover, one would expect the $tRNA^{Tyr}$ probe to hybridize to defective or structurally altered $tRNA^{Tyr}$ molecules, leading to the visualization of more than one band on the membrane. Alternatively, lower expression of $tRNA^{Tyr}$ could be explained by increased turnover of the mutant $tRNA^{Tyr}$, which based on our results, is the more likely explanation.

Functional and morphologic analyses of mitochondria

To assess the mitochondrial ATP-production rate (MAPR) and respiratory chain enzyme activity, muscle tissue was collected from five affected and five healthy dogs. In affected dogs, decreased MAPR was observed in isolated mitochondria, compared to controls. Respiratory chain enzyme activity was significantly decreased in complex I (-63 %), complex I-III (-53 %) and complex IV (-59 %) in affected dogs compared to controls. There was no altered activity of complex II between affected and control dogs. Interestingly, complex II is encoded exclusively by nDNA. Measurements of MAPR and respiratory chain enzyme activity show consistent results, indicating deficiency of complex I and IV, which are dependent of subunits encoded by the mtDNA. Although the clinical examination did not show any signs of a muscle disorder, the MAPR and respiratory chain enzyme activity revealed severely affected mitochondrial function in affected dogs. Similar mitochondrial dysfunction has been documented in humans and mice with mitochondrial disorders (Bindoff *et al.*, 1991; von Döbeln *et al.*, 1993; Vielhaber *et al.*, 2002; Trifunovic *et al.*, 2004). Measurements of citrate synthase activity in muscle showed elevated activity (+47 %) in cases, indicating increased mitochondrial proliferation, which has been observed in tissues with respiratory chain enzyme deficiencies. It is believed that the increased mitochondrial proliferation is due to the inadequate energy level, thus it may be a compensatory reaction in affected tissues. One of the affected dogs was indistinguishable from the controls in the experiments described above.

Morphological changes were observed in muscle biopsies utilized in the MAPR and respiratory chain enzyme activity experiments. Four of the five affected dogs showed reduced COX (cytochrome c) activity as compared to SDH (succinate dehydrogenase). Muscle fibres with absent COX-activity and red-ragged fibres are observed in humans with large-scale deletions and *tRNA* mutations in mtDNA (Bindoff *et al.*, 1991; Vielhaber *et al.*, 2002; Bourgeois & Tarnopolsky, 2004). Surprisingly, red-ragged fibres were not observed in SAN dogs. This could be due to the fact that SAN cases do not display any signs of muscle disease. Alternatively, this finding could be explained by the different muscle composition and higher mitochondrial density in dogs compared to humans (Wakshlag *et al.*, 2004). Interestingly, mitochondrial paracrystaline inclusions were identified in the affected dog that was indistinguishable from the controls by MAPR and respiratory chain enzyme measurements. Paracrystaline inclusions are a common finding in adult humans presenting mitochondrial myopathies.

Concluding remarks of SAN

The data presented in this study provides conclusive evidence that SAN is a maternally inherited mitochondrial disorder caused by a one bp deletion in the $tRNA^{Tyr}$ gene (**$\Delta T5304$**). Furthermore, the mutation does not cause a clinical phenotype unless the mutant load is very high. Mutations of the mitochondrial $tRNA^{Tyr}$ gene have been reported in humans, although they involve other positions and the patients present clinical symptoms other than those observed in SAN affected dogs (Pulkes *et al.*, 2000; Raffelsberger *et al.*, 2001; Sahashi *et al.*, 2001; Scaglia *et al.*, 2003). Although mouse models of several mitochondrial disorders have been established, they do not always present satisfying models for the corresponding spontaneously occurring mitochondrial disease in humans. While both harbour the corresponding mitochondrial mutations, the clinical symptoms and the course of the disease may differ considerably. However, this scenario of the same mutation and different clinical symptoms is not uncommon among human patients. Intriguingly, SAN affected dogs present strikingly similar clinical symptoms. Finally, we believe that SAN in Golden Retrievers may constitute an interesting model for studying human mtDNA disorders.

Undoubtedly, this study, as well as related studies describing the clinical aspects of SAN (Jaderlund *et al.*, 2007; Jaderlund, 2009), has had a great impact on the Swedish Golden Retriever population. Following the identification of the causative mutation, we also designed a genetic test for SAN and made this available to the public. We recommended genetic testing of suspected SAN cases as well as potential breeding female relatives.

46

So far, blood from 15 Golden Retrievers has been tested. Four unrelated dogs were identified as non-carriers, nine maternally related dogs were carriers and two dogs belonging to the SAN family showed a homoplasmy of the wt allele. Hence, we recommend testing offspring descending from these two females, since we cannot exclude the possibility that low levels of the mutant allele are present in other tissues, such as egg cells.

Future prospects

Challenges with mitochondrial disorders and future of SAN

The most challenging and time consuming part of this project was to find a quantification method which satisfied our needs for accuracy, resolution and which showed stable and consistent measurement. Although we found a method (qOLA) that met these terms, we are still puzzled by the overlapping groups of affected dogs and their close relatives. Imperfect genotype-phenotype correlations seem to be one of the greater challenges that mitochondrial geneticists have to solve. Thus, many questions remain to be answered, for example: What part does the nuclear genetic background play in the clinical penetrance of SAN? Is there a reliable way to identify environmental factors?

The primary tissues used in this study were blood and muscle, however SAN is manifested in the CNS, primarily in sensory neurons, and at present the biological mechanism in sensory neurons which give rise to this particular disease is not understood. Currently the only test we have for screening for SAN involves assessing the mutant load in blood or muscle tissue as a proxy for CNS affection. Clearly more investigation into unravelling the biological mechanism underpinning this disease of the CNS, primarily sensory neurons, is required.

There may be two possible ways to further assess the functional effect of this $tRNA^{Tyr}$ deletion. One option would be to further characterize the mutant $tRNA^{Tyr}$ molecule *in vitro*, for example by studying the codon-anticodon binding, or the amino acylation process by which amino acids are attached to the 3' end of the *tRNA* molecule. An alternative or complementary approach would be to generate a cybrid cell line harbouring this specific deletion, allowing manipulation in an *in vivo* system.

Even though we have not identified the corresponding disorder in humans, we still believe that SAN offers a model for studying spontaneously occurring mitochondrial disorders, and possibly may be utilized for testing different therapeutic approaches which could benefit both dogs and humans.

Populärvetenskaplig sammanfattning

Kartläggning av gener associerade med en egenskap (vit päls och hudfärg) och en sjukdom (sensorisk ataktisk neuropati) understryker hundens potential som modellorganism

Bakgrund

Sedan urminnes tider, har människan försökt att förstå hur olika egenskaper så som exempelvis kroppslängd kan förklaras och påverkas. Speciellt intresse har riktats till att förstå hur det kan komma sig att särskilda sjukdomar enbart drabbar vissa individer. Ibland kan detta förklaras med olika miljöfaktorer (t.ex. osund livsstil), men allt oftare ligger svaret i arvsmassan.

När det av olika anledningar är omöjligt att hitta orsaken bakom en viss egenskap eller sjukdom hos människor, har ofta djur nyttjats, s.k. modellorganismer. De vanligaste modellorganismerna inkluderar zebrafisk, rundmask, bananfluga och mus. På senare tid har vi även upptäckt de stora möjligheter som finns med att nyttja mångfalden hos våra husdjur för att förstå hur arvsmassan i samverkan med miljöfaktorer påverkar egenskaper. Följaktligen har såväl produktionsegenskaper och sjukdomar med relevans för människan kartlagts genetiskt hos våra husdjur.

I denna avhandling beskrivs hur gener som påverkar olika egenskaper och sjukdomar kan kartläggas genom att nyttja hunden som modelldjur i genetiska studier.

Hunden som modelldjur

Hunden, människans bästa vän, har en unik populationshistoria som utgörs av minst två s.k. flaskhalsar, som innebär att en population drastiskt reduceras i antal under en period för att sedan öka igen (Figur 4). Den första flaskhalsen inträffade när tamhunden tämjdes från vargen ca 15,000–31,000 år sedan. Denna population bestod av individer som parade sig slumpvis och

kunde se väldigt olika ut. Avelsbasen var bred och variationen var stor inom hundpopulationen. Aveln var inte i någon större utsträckning påverkad av människan. Successivt började människor att välja ut hundar som uppvisade olika önskade egenskaper, till en början beteenden som vakt, jakt och vallning, och på senare tid även utseenden, t.ex. pälsfärg och kroppsstorlek.

För ungefär tvåhundra år sedan påbörjade människan ett av de mest omfattande avelsexperimenten: renrasavel hos hund. Detta innebar att hundarna delades upp i enskilda raser och strikta avelskriterier definierades för varje ras. Detta var den andra flaskhalsen. Mer konkret innebar det att de flesta raserna baserades på väldigt få individer och variationen inom varje ras minskade drastiskt. Däremot uppvisar hela hundpopulationen mer variation än något annat husdjur, med fler än 400 raser registrerade.

Selektiv renrasavel och genetisk drift (slumpmässiga förändringar av allelfrekvenser), till följd av att en liten andel av populationen nyttjas i aveln, har gett upphov till att förekomsten av vissa oönskade egenskaper och sjukdomar har ökat inom olika raser. Detta omfattar olika beteenden t.ex. aggressivitet och även sjukdomar så som cancer, diabetes och epilepsi.

Hundens unika populationshistoria samt dess sjukdomsbild, som ofta liknar människans, gör hunden till ett intressant modelldjur som kan nyttjas för att hitta sjukdomsorsakande gener. Dessa studier skulle på sikt kunna hjälpa hundar såväl som människor.

Delarbete 1 och 3: Kartläggning av *MITF* genen som ger upphov till vit pälsfärg hos hund bekräftar att vår två-stegs princip fungerar

I det första delarbetet av denna avhandling beskrivs en kartläggningsmetod för gener hos hundar som påverkar olika egenskaper och sjukdomar. Orsaken bakom vit pälsfärg och ljus hudfärg hos hund, som är en relativt enkel egenskap att klassificera, har kartlagts med hjälp av en två-stegs princip. I det första steget jämfördes arvsmassan hos vita och enfärgade mörka boxrar, och en omfattande region i arvsmassan som är kopplad till egenskapen identifierades. Den begränsade variationen inom en ras nyttjades, och därigenom var det tillräckligt att använda endast ett litet antal hundar och relativt få genetiska markörer. I det andra steget analyserades den associerade region hos andra raser som uppvisar samma färgteckning, i denna studie användes bullterrier. När fler raser jämförs kan vi dra nytta av den omfattande (genetiska) variationen mellan raser. Härigenom lyckades vi minska regionen i arvsmassan som ger upphov till vit pälsfärg 10-faldigt och genen som orsakar vitfärg kunde identifieras som den s.k. *microphtalmia-associated transcription factor* (*MITF*) genen. Analyser av *MITF* genen avslöjade två potentiella kandidatmutationer som utvärderades för att fastställa hur

mutationerna påverkar uttrycket av *MITF* genen. Denna gen är även förknippad med bland annat avvikande pigmentering och dövhet hos mus och människa. Dövhet har en högre förekomst hos hundar med vit päls och ljus hud.

Delarbete 2: Sensorisk ataktisk neuropati hos golden retriever orsakas av en mutation i den mitokondriella arvsmassan

Den andra delen (delarbete 2) i avhandlingen beskriver en av de första mitokondriella sjukdomarna hos hund. Mitokondrier är organeller som bland annat producerar energi (ATP) för våra celler. Mitokondrierna har även en egen arvsmassa som nedärvs på mödernet, dvs. från mor till avkomma. Hos människa har fler än 550 sjukdomsorsakande mutationer i den mitokondriella arvsmassan identifierats och förekomsten av mitokondriella sjukdomar har uppskattats till ca 1/5000, vilket anses vara relativt vanligt.

  Sensorisk ataktisk neuropati (SAN) är en neurologisk sjukdom som drabbar hundar som tillhör en särskild familj av rasen golden retriever. Genom att analysera familjeträd fastställdes att samtliga fall härstammar (på mödernet) från en tik som levde på 1970-talet. Analys av mitokondriell arvsmassa avslöjade en potentiell kandidatmutation som utvärderades, varefter det fastställdes att den är sjukdomsorsakande. Omfattande analyser av mitokondrier bekräftade att sjukdomen har sitt ursprung i mitokondrierna. I samband med att SAN-studien publicerades utvecklades ett genetiskt test som möjliggör att SAN nu kan avlas bort från golden retriever populationen.

Slutsats

Dessa två studier som beskriver kartläggningen av en egenskap och en sjukdom, med ursprung i två olika arvsmassor, visar utan tvivel styrkan med genetiska studier på hund. Därmed har en ny roll för människans bästa vän etablerats som skapar goda förutsättningar att på sikt även kunna avla friskare hundar.

References

Altshuler, D., Daly, M.J. & Lander, E.S. (2008). Genetic mapping in human disease. *Science* 322(5903), 881-8.

Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Bonnen, P.E., de Bakker, P.I., Deloukas, P., Gabriel, S.B., Gwilliam, R., Hunt, S., Inouye, M., Jia, X., Palotie, A., Parkin, M., Whittaker, P., Chang, K., Hawes, A., Lewis, L.R., Ren, Y., Wheeler, D., Muzny, D.M., Barnes, C., Darvishi, K., Hurles, M., Korn, J.M., Kristiansson, K., Lee, C., McCarrol, S.A., Nemesh, J., Keinan, A., Montgomery, S.B., Pollack, S., Price, A.L., Soranzo, N., Gonzaga-Jauregui, C., Anttila, V., Brodeur, W., Daly, M.J., Leslie, S., McVean, G., Moutsianas, L., Nguyen, H., Zhang, Q., Ghori, M.J., McGinnis, R., McLaren, W., Takeuchi, F., Grossman, S.R., Shlyakhter, I., Hostetter, E.B., Sabeti, P.C., Adebamowo, C.A., Foster, M.W., Gordon, D.R., Licinio, J., Manca, M.C., Marshall, P.A., Matsuda, I., Ngare, D., Wang, V.O., Reddy, D., Rotimi, C.N., Royal, C.D., Sharp, R.R., Zeng, C., Brooks, L.D. & McEwen, J.E. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* 467(7311), 52-8.

Andersson, L. (2009). Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. *Genetica* 136(2), 341-9.

Arnaud, P., Goubely, C., Pelissier, T. & Deragon, J.M. (2000). SINE retroposons can be used in vivo as nucleation centers for de novo methylation. *Mol Cell Biol* 20(10), 3434-41.

Awano, T., Johnson, G.S., Wade, C.M., Katz, M.L., Johnson, G.C., Taylor, J.F., Perloski, M., Biagi, T., Baranowska, I., Long, S., March, P.A., Olby, N.J., Shelton, G.D., Khan, S., O'Brien, D.P., Lindblad-Toh, K. & Coates, J.R. (2009). Genome-wide association analysis reveals a SOD1 mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis. *Proc Natl Acad Sci U S A* 106(8), 2794-9.

Bekris, L.M., Yu, C.E., Bird, T.D. & Tsuang, D.W. (2010). Genetics of Alzheimer disease. *Journal of geriatric psychiatry and neurology* 23(4), 213-27.

Bindoff, L.A., Desnuelle, C., Birch-Machin, M.A., Pellissier, J.F., Serratrice, G., Dravet, C., Bureau, M., Howell, N. & Turnbull, D.M. (1991). Multiple defects of the mitochondrial respiratory chain in a mitochondrial encephalopathy (MERRF): a clinical, biochemical nad molecular study. *J Neurol Sci* 102(1), 17-24.

Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32(3), 314-31.

Bourgeois, J.M. & Tarnopolsky, M.A. (2004). Pathology of skeletal muscle in mitochondrial disorders. *Mitochondrion* 4(5-6), 441-452.

Cao, L., Shitara, H., Horii, T., Nagao, Y., Imai, H., Abe, K., Hara, T., Hayashi, J. & Yonekawa, H. (2007). The mitochondrial bottleneck occurs without reduction of mtDNA content in female mouse germ cells. *Nat Genet* 39(3), 386-90.

Collins, F.S., Guyer, M.S. & Charkravarti, A. (1997). Variations on a theme: cataloging human DNA sequence variation. *Science* 278(5343), 1580-1.

Consortium, T.I.H. (2005). A haplotype map of the human genome. *Nature* 437(7063), 1299-320.

Consortium, T.W.T.C.C. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145), 661-78.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. & Weissenbach, J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380(6570), 152-4.

DiMauro, S. & Davidzon, G. (2005). Mitochondrial DNA and disease. *Ann Med* 37, 222-232.

DiMauro, S. & Schon, E.A. (2001). Mitochondrial DNA mutations in human disease. *Am J Med Genet* (106), 18.

DiMauro, S. & Schon, E.A. (2003). Mitochondrial respiratory-chain diseases. *N Engl J Med* 348, 2656-2668.

Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T.P., Bowden, D.W., Smith, D.R., Lander, E.S. & et al. (1987). A genetic linkage map of the human genome. *Cell* 51(2), 319-37.

Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q.,

Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Sun, W., Wang, H., Wang, Y., Xiong, X., Xu, L., Waye, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., Song, Y.Q., Tam, P.K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., de Bakker, P.I., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Schaffner, S.F., Sham, P.C., Varilly, P., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Yakub, I., Birren, B.W., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archeveque, P., Bellemare, G., Saeki, K., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R. & Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449(7164), 851–61.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart,

M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J. & Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science* 296(5576), 2225-9.

Graff, C., Bui, T.H. & Larsson, N.G. (2002). Mitochondrial disease. *Best Pract Res Clin Obstet Gynaecol* 16(5), 715-728.

Greaves, L.C. & Taylor, R.W. (2006). Mitchondrial DNA mutations in human disease. *IUBMB Life* 58(3), 143-151.

Hirschhorn, J.N. & Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6(2), 95-108.

Holt, I.J., Harding, A.E. & Morgan-Hughes, J.A. (1988). Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* 331(6158), 717-9.

Illa, I., Rojas, R., Gallardo, E., Serrano, C. & Graus, F. (2001). Chronic idiopathic sensory ataxic neuropathy: immunological aspects of a series of 17 patients. *Rev Neurol (Paris)* 157(5), 517-22.

Jaderlund, K.H. (2009). Mitochondrially inherited sensory ataxic neuropathy in golden retriever dogs. Phenotype, clinical course and genotype of a novel neurological syndrome. *Doctoral Thesis, Swedish University of Agricultural Sciences* No 2009:73.

Jaderlund, K.H., Orvind, E., Johnsson, E., Matiasek, K., Hahn, C.N., Malm, S. & Hedhammar, A. (2007). A neurologic syndrome in Golden Retrievers presenting as a sensory ataxic neuropathy. *J Vet Intern Med* 21(6), 1307-15.

Jenuth, J.P., Peterson, A.C., Fu, K. & Shoubridge, E.A. (1996). Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nat Genet* 14(2), 146-51.

Jonasdottir, T.J., Mellersh, C.S., Moe, L., Heggebo, R., Gamlem, H., Ostrander, E.A. & Lingaas, F. (2000). Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs. *Proc Natl Acad Sci U S A* 97(8), 4132-7.

Karlsson, E.K. & Lindblad-Toh, K. (2008). Leader of the pack: gene mapping in dogs and other model organisms. *Nature reviews. Genetics* 9(9), 713-25.

King, D.C., Taylor, J., Elnitski, L., Chiaromonte, F., Miller, W. & Hardison, R.C. (2005). Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. *Genome Res* 15(8), 1051-60.

Kolbe, D., Taylor, J., Elnitski, L., Eswara, P., Li, J., Miller, W., Hardison, R. & Chiaromonte, F. (2004). Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat. *Genome Res* 14(4), 700-7.

Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 22(2), 139-44.

Lander, E.S. (1996). The new genomics: global views of biology. *Science* 274(5287), 536-9.

Lander, E.S. & Schork, N.J. (1994). Genetic dissection of complex traits. *Science* 265(5181), 2037-48.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., Ferreira, T., Wood, A.R., Weyant, R.J., Segre, A.V., Speliotes, E.K., Wheeler, E., Soranzo, N., Park, J.H., Yang, J., Gudbjartsson, D., Heard-Costa, N.L., Randall, J.C., Qi, L., Vernon Smith, A., Magi, R., Pastinen, T., Liang, L., Heid, I.M., Luan, J., Thorleifsson, G., Winkler, T.W., Goddard, M.E., Sin Lo, K., Palmer, C., Workalemahu, T., Aulchenko, Y.S., Johansson, A., Zillikens, M.C., Feitosa, M.F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M., Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N.L., Hayward, C., Hottenga, J.J., Jacobs, K.B., Knowles, J.W., Kutalik, Z., Monda, K.L., Polasek, O., Preuss, M., Rayner, N.W., Robertson, N.R., Steinthorsdottir, V., Tyrer, J.P., Voight, B.F., Wiklund, F., Xu, J., Zhao, J.H., Nyholt, D.R., Pellikka, N., Perola, M., Perry, J.R., Surakka, I., Tammesoo, M.L., Altmaier, E.L., Amin, N., Aspelund, T., Bhangale, T., Boucher, G., Chasman, D.I., Chen, C., Coin, L., Cooper, M.N., Dixon, A.L., Gibson, Q., Grundberg, E., Hao, K., Juhani Junttila, M., Kaplan, L.M., Kettunen, J., Konig, I.R., Kwan, T., Lawrence, R.W., Levinson, D.F., Lorentzon, M., McKnight, B., Morris, A.P., Muller, M., Suh Ngwa, J., Purcell, S., Rafelt, S., Salem, R.M., Salvi, E., Sanna, S., Shi, J., Sovio, U., Thompson, J.R., Turchin, M.C., Vandenput, L., Verlaan, D.J., Vitart, V., White, C.C., Ziegler, A., Almgren, P., Balmforth, A.J., Campbell, H., Citterio, L., De Grandi, A., Dominiczak, A., Duan, J., Elliott, P., Elosua, R., Eriksson, J.G., Freimer, N.B., Geus, E.J., Glorioso, N., Haiqing, S., Hartikainen, A.L., Havulinna, A.S., Hicks, A.A., Hui, J., Igl, W., Illig, T., Jula, A., Kajantie, E., Kilpelainen, T.O., Koiranen, M., Kolcic, I., Koskinen, S., Kovacs, P., Laitinen, J., Liu, J., Lokki, M.L., Marusic, A., Maschio, A., Meitinger, T., Mulas, A., Pare, G., Parker, A.N., Peden, J.F., Petersmann, A., Pichler, I., Pietilainen, K.H., Pouta, A., Ridderstrale, M., Rotter, J.I., Sambrook, J.G., Sanders, A.R., Schmidt, C.O., Sinisalo, J., Smit, J.H., Stringham, H.M., Bragi Walters, G., Widen, E., Wild, S.H., Willemsen, G., Zagato, L., Zgaga, L., Zitting, P., Alavere, H., Farrall, M., McArdle, W.L., Nelis, M., Peters, M.J., Ripatti, S., van Meurs, J.B., Aben, K.K., Ardlie, K.G., Beckmann, J.S., Beilby, J.P., Bergman, R.N., Bergmann, S., Collins, F.S., Cusi, D., den Heijer, M., Eiriksdottir, G., Gejman, P.V., Hall, A.S., Hamsten, A., Huikuri, H.V., Iribarren, C., Kahonen, M., Kaprio, J., Kathiresan, S., Kiemeney, L., Kocher, T., Launer, L.J., Lehtimaki, T.,

Melander, O., Mosley, T.H., Jr., Musk, A.W., Nieminen, M.S., O'Donnell, C.J., Ohlsson, C., Oostra, B., Palmer, L.J., Raitakari, O., Ridker, P.M., Rioux, J.D., Rissanen, A., Rivolta, C., Schunkert, H., Shuldiner, A.R., Siscovick, D.S., Stumvoll, M., Tonjes, A., Tuomilehto, J., van Ommen, G.J., Viikari, J., Heath, A.C., Martin, N.G., Montgomery, G.W., Province, M.A., Kayser, M., Arnold, A.M., Atwood, L.D., Boerwinkle, E., Chanock, S.J., Deloukas, P., Gieger, C., Gronberg, H., Hall, P., Hattersley, A.T., Hengstenberg, C., Hoffman, W., Lathrop, G.M., Salomaa, V., Schreiber, S., Uda, M., Waterworth, D., Wright, A.F., Assimes, T.L., Barroso, I., Hofman, A., Mohlke, K.L., Boomsma, D.I., Caulfield, M.J., Cupples, L.A., Erdmann, J., Fox, C.S., Gudnason, V., Gyllensten, U., Harris, T.B., Hayes, R.B., Jarvelin, M.R., Mooser, V., Munroe, P.B., Ouwehand, W.H., Penninx, B.W., Pramstaller, P.P., Quertermous, T., Rudan, I., Samani, N.J., Spector, T.D., Volzke, H., Watkins, H., Wilson, J.F., Groop, L.C., Haritunians, T., Hu, F.B., Kaplan, R.C., Metspalu, A., North, K.E., Schlessinger, D., Wareham, N.J., Hunter, D.J., O'Connell, J.R., Strachan, D.P., Wichmann, H.E., Borecki, I.B., van Duijn, C.M., Schadt, E.E., Thorsteinsdottir, U., Peltonen, L., Uitterlinden, A.G., Visscher, P.M., Chatterjee, N., Loos, R.J., Boehnke, M., McCarthy, M.I., Ingelsson, E., Lindgren, C.M., Abecasis, G.R., Stefansson, K., Frayling, T.M. & Hirschhorn, J.N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467(7317), 832-8.

Levy, C., Khaled, M. & Fischer, D.E. (2006). MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol Med* 12(9), 406-414.

Li, F.Y., Cuddon, P.A., Song, J., Wood, S.L., Patterson, J.S., Sheldon, G.D. & Duncan, I.D. (2006). Canine spongiform leukoencephalomyelopathy is associated with a missense mutation in cytochrome b. *Neurobiol Dis* 21(1), 35–42.

Lin, J.Y. & Fisher, D.E. (2007). Melanocte biology and skin pigmentation. *Nature* 445(7130), 843–850.

Lin, L., Faraco, J., Li, R., Kadotani, H., Rogers, W., Lin, X., Qiu, X., de Jong, P.J., Nishino, S. & Mignot, E. (1999). The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* 98(3), 365–76.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., 3rd, Zody, M.C., Mauceli, E., Xie, X., Breen, M., Wayne, R.K., Ostrander, E.A., Ponting, C.P., Galibert, F., Smith, D.R., DeJong, P.J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C.W., Cook, A., Cuff, J., Daly, M.J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C.,

Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K.P., Parker, H.G., Pollinger, J.P., Searle, S.M., Sutter, N.B., 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., 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, A.C., Kisner, P., Kumar, M., Lance, K., Landers, T., Lara, M., Lee, W., Leger, J.P., 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., 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, E.S. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069), 803-19.

Lingaas, F., Comstock, K.E., Kirkness, E.F., Sorensen, A., Aarskaug, T., Hitte, C., Nickerson, M.L., Moe, L., Schmidt, L.S., Thomas, R., Breen, M., Galibert, F., Zbar, B. & Ostrander, E.A. (2003). A mutation in the canine BHD gene is associated with hereditary

multifocal renal cystadenocarcinoma and nodular dermatofibrosis in the German Shepherd dog. *Hum Mol Genet* 12(23), 3043-53.

Little, C.C. (1957). The inheritance of coat colour in the dogs. *Comstock Publishing Associates, Cornell University Press, Ithaca, NY.*

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F., McCarroll, S.A. & Visscher, P.M. (2009). Finding the missing heritability of complex diseases. *Nature* 461(7265), 747-53.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A.E. & Wingender, E. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 34(Database issue), D108-10.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P. & Hirschhorn, J.N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9(5), 356-69.

Mellersh, C.S., Langston, A.A., Acland, G.M., Fleming, M.A., Ray, K., Wiegand, N.A., Francisco, L.V., Gibbs, M., Aguirre, G.D. & Ostrander, E.A. (1997). A linkage map of the canine genome. *Genomics* 46(3), 326-36.

Mignot, E., Wang, C., Rattazzi, C., Gaiser, C., Lovett, M., Guilleminault, C., Dement, W.C. & Grumet, F.C. (1991). Genetic linkage of autosomal recessive canine narcolepsy with a mu immunoglobulin heavy-chain switch-like segment. *Proc Natl Acad Sci U S A* 88(8), 3475-8.

Miller, W., Rosenbloom, K., Hardison, R.C., Hou, M., Taylor, J., Raney, B., Burhans, R., King, D.C., Baertsch, R., Blankenberg, D., Kosakovsky Pond, S.L., Nekrutenko, A., Giardine, B., Harris, R.S., Tyekucheva, S., Diekhans, M., Pringle, T.H., Murphy, W.J., Lesk, A., Weinstock, G.M., Lindblad-Toh, K., Gibbs, R.A., Lander, E.S., Siepel, A., Haussler, D. & Kent, W.J. (2007). 28-way vertebrate alignment and conservation track in the UCSC Genome Browser. *Genome Res* 17(12), 1797-808.

Munoz-Sanjuan, I. & Bates, G.P. (2011). The importance of integrating basic and clinical research toward the development of new therapies for Huntington disease. *J Clin Invest* 121(2), 476-83.

Naviaux, R.K. (2000). Mitochondrial DNA disorders. *Eur J Pediatr* 159 Suppl 3(Dec), S219.

Olsson, M., Meadows, J.R., Truve, K., Rosengren Pielberg, G., Puppo, F., Mauceli, E., Quilez, J., Tonomura, N., Zanna, G., Docampo, M.J., Bassols, A., Avery, A.C., Karlsson, E.K., Thomas, A., Kastner, D.L., Bongcam-Rudloff, E., Webster, M.T., Sanchez, A., Hedhammar, A., Remmers, E.F., Andersson, L., Ferrer, L., Tintle, L. & Lindblad-Toh, K. (2011). 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* 7(3), e1001332.

Ostrander, E.A., Galibert, F. & Patterson, D.F. (2000). Canine genetics comes of age. *Trends Genet* 16(3), 117-24.

Park, C.B. & Larsson, N.G. (2011). Mitochondrial DNA mutations in disease and aging. *J Cell Biol* 193(5), 809-18.

Patterson, D.F. (2000). Companion animal medicine in the age of medical genetics. *J Vet Intern Med* 14(1), 1-9.

Peltonen, L., Palotie, A. & Lange, K. (2000). Use of population isolates for mapping complex traits. *Nat Rev Genet* 1(3), 182-90.

Peltonen, L., Perola, M., Naukkarinen, J. & Palotie, A. (2006). Lessons from studying monogenic disease for common disease. *Hum Mol Genet* 15 Spec No 1, R67-74.

Price, A.L., Zaitlen, N.A., Reich, D. & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nat Rev Genet* 11(7), 459-63.

Pulkes, T., Siddiqui, A., Morgan-Hughes, J.A. & Hanna, M.G. (2000). A novel mutation in the mitochondrial tRNA Tyr gene associated with exercise intolerance. *Neurology* 55, 1210-1212.

Raffelsberger, T., Rossmanith, W., Thaller-Antlanger, H. & Bittner, R.E. (2001). CPEO associated with a single nucleotide deletion in the mitochondrial tRNA Tyr gene. *Neurology* 57, 2298-2301.

Risch, N. & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* 273(5281), 1516-7.

Ronaghi, M. & Elahi, E. (2002). Discovery of single nucleotide polymorphisms and mutations by pyrosequencing. *Comp Funct Genomics* 3(1), 51-6.

Rosengren Pielberg, G., Golovko, A., Sundstrom, E., Curik, I., Lennartsson, J., Seltenhammer, M.H., Druml, T., Binns, M., Fitzsimmons, C., Lindgren, G., Sandberg, K., Baumung, R., Vetterlein, M., Stromberg, S., Grabherr, M., Wade, C., Lindblad-Toh, K., Ponten, F., Heldin, C.H., Solkner, J. & Andersson, L. (2008). A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nat Genet* 40(8), 1004-9.

Rubin, C.J., Brandstrom, H., Wright, D., Kerje, S., Gunnarsson, U., Schutz, K., Fredriksson, R., Jensen, P., Andersson, L., Ohlsson, C., Mallmin, H., Larsson, S. & Kindmark, A. (2007). Quantitative trait

loci for BMD and bone strength in an intercross between domestic and wildtype chickens. *J Bone Miner Res* 22(3), 375-84.

Ruiz-Pesini, E., Lott, M.T., Procaccio, V., Poole, J.C., Brandon, M.C., Mishmar, D., Yi, C., Kreuziger, J., Baldi, P. & Wallace, D.C. (2007). An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res* 35(Database issue), D823-8.

Sahashi, K., Yoneda, M., Ohno, K., Tanaka, M., Ibi, T. & Sahashi, K. (2001). Functional characteristics of mitochondrial tRNATyr mutation (5877G->A) associated with familial chronic progressive external ophthalmoplegia. *J Med Genet* 38(10), 703-705.

Scaglia, F., Vogel, H., Hawkins, E.P., Vladutiu, G.D., Liu, L.L. & Wong, L.J. (2003). Novel homoplasmic mutation in the mitochondrial tRNA Tyr gene associated with atypical mitochondrial cytopathy presenting with focal segmental glomerulosclerosis. *Am J Med Genet A* 123A, 172-178.

Schaefer, A.M., Taylor, R.W., Turnbull, D.M. & Chinnery, P.F. (2004). The epidemiology of mitochondrial disorders--past, present and future. *Biochim Biophys Acta* 1659(2-3), 115-20.

Schuster, S.C. (2008). Next-generation sequencing transforms today's biology. *Nat Methods* 5(1), 16-8.

Seo, B.Y., Park, E.W., Ahn, S.J., Lee, S.H., Kim, J.H., Im, H.T., Lee, J.H., Cho, I.C., Kong, I.K. & Jeon, J.T. (2007). An accurate method for quantifying and analyzing copy number variation in porcine KIT by an oligonucleotide ligation assay. *BMC Genet* 8, 81.

Service, S., DeYoung, J., Karayiorgou, M., Roos, J.L., Pretorious, H., Bedoya, G., Ospina, J., Ruiz-Linares, A., Macedo, A., Palha, J.A., Heutink, P., Aulchenko, Y., Oostra, B., van Duijn, C., Jarvelin, M.R., Varilo, T., Peddle, L., Rahman, P., Piras, G., Monne, M., Murray, S., Galver, L., Peltonen, L., Sabatti, C., Collins, A. & Freimer, N. (2006). Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet* 38(5), 556-60.

Smith, S.D., Kelley, P.M., Kenyon, J.B. & Hoover, D. (2000). Tietz syndrome (hypopigmentation/deafness) caused by mutation of MITF. *J Med Genet* 37(6), 446-8.

Steingrimsson, E., Copeland, N.G. & Jenkins, N.A. (2004). Melanocytes and the microphthalmia transcription factor network. *Annu Rev Genet* 38, 365-411.

Strain, G.M. (2004). Deafness prevalence and pigmentation and gender associations in dog breeds at risk. *Veterinary journal* 167(1), 23-32.

Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C. & Schatten, G. (1999). Ubiquitin tag for sperm mitochondria. *Nature* 402(6760), 371-2.

Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C. & Schatten, G. (2000). Ubiquitinated sperm mitochondria,

selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol Reprod* 63(2), 582-90.

Sutter, N.B., Eberle, M.A., Parker, H.G., Pullar, B.J., Kirkness, E.F., Kruglyak, L. & Ostrander, E.A. (2004). Extensive and breed-specific linkage disequilibrium in Canis familiaris. *Genome Res* 14(12), 2388-96.

Sutter, N.B. & Ostrander, E.A. (2004). Dog star rising: the canine genetic system. *Nat Rev Genet* 5(12), 900-10.

Tassabehji, M., Newton, V.E. & Read, A.P. (1994). Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. *Nat Genet* 8(3), 251-5.

Taylor, R.W., Giordano, C., Davidson, M.M., d'Amati, G., Bain, H., Hayes, C.M., Leonard, H., Barron, M.J., Casali, C., Santorelli, F.M., Hirano, M., Lightowlers, R.N., DiMauro, S. & Turnbull, D.M. (2003). A homoplasmic mitochondrial transfer ribonucleic acid mutation as a cause of maternally inherited hypertrophic cardiomyopathy. *J Am Coll Cardiol* 41(10), 1786-1796.

Taylor, R.W. & Turnbull, D.M. (2005). Mitochondrial DNA mutations in human disease. *Nat Rev Genet* 6(May), 389-402.

Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J.N., Rovio, A.T., Bruder, C.E., Bohlooly, Y.M., Gidlof, S., Oldfors, A., Wibom, R., Tornell, J., Jacobs, H.T. & Larsson, N.G. (2004). Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* 429(6990), 417-423.

van De Sluis, B., Rothuizen, J., Pearson, P.L., van Oost, B.A. & Wijmenga, C. (2002). Identification of a new copper metabolism gene by positional cloning in a purebred dog population. *Hum Mol Genet* 11(2), 165-73.

Van Laere, A.S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A.L., Haley, C.S., Buys, N., Tally, M., Andersson, G., Georges, M. & Andersson, L. (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* 425(6960), 832-6.

Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Rosengren Pielberg, G., Sigurdsson, S., Fall, T., Seppälä, E., Hansen, M.S.T., Lawley, C.T., Karlsson, E.K., consortium, T.L., Bannasch, D., Vilà, C., Lohi, H., Galibert, F., Fredholm, M., Hedhammar, Å., André, C., Lindblad-Toh, K., Hitte, C. & Webster, M.T. (2011). Identification of loci governing common and rare phenotypic traits in dog breeds using selection mapping. *PLoS Genet*, Accepted.

Vielhaber, S., Varlamov, D.A., Kudina, T.A., Schröder, R., Kappes-Horn, K., Elger, C.E., Seibel, M., Seibel, P. & Kunz, W.S. (2002). Expression pattern of mitochondrial respiratory chain enzymes in skeletal muscle of patients harboring the A3243G point mutation or

large-scale deletions of mitochondrial DNA. *J Neuropathol Exp Neurol* 61(10), 885-895.

von Döbeln, U., Wibom, R., Ahlman, H., Nennesmo, I., Nyctelius, H., Hultman, E. & Hagenfeldt, L. (1993). Fatal neonatal lactic acidosis with respiratory insufficiency due to complex I and IV deficiency. *Acta Paediatr.* 82(12), 1079-1081.

Vonholdt, B.M., Pollinger, J.P., Lohmueller, K.E., Han, E., Parker, H.G., Quignon, P., Degenhardt, J.D., Boyko, A.R., Earl, D.A., Auton, A., Reynolds, A., Bryc, K., Brisbin, A., Knowles, J.C., Mosher, D.S., Spady, T.C., Elkahloun, A., Geffen, E., Pilot, M., Jedrzejewski, W., Greco, C., Randi, E., Bannasch, D., Wilton, A., Shearman, J., Musiani, M., Cargill, M., Jones, P.G., Qian, Z., Huang, W., Ding, Z.L., Zhang, Y.P., Bustamante, C.D., Ostrander, E.A., Novembre, J. & Wayne, R.K. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464(7290), 898-902.

Wakshlag, J.J., Cooper, B.J., Wakshlag, R.R., Kallfelz, F.A., Barr, S.C., Nydam, D.V. & Dimauro, S. (2004). Biochemical evaluation of mitochondrial respiratory chain enzymes in canine skeletal muscle. *Am J Vet Res* 65(4), 480-484.

Wallace, D.C. (2001). Mouse models for mitochondrial disease. *Am J Med Genet* 106, 71.

Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A., Schurr, T.G., Lezza, A.M., Elsas, L.J., 2nd & Nikoskelainen, E.K. (1988). Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science* 242(4884), 1427-30.

Wibom, R., Hagenfeldt, L. & von Döbeln, U. (2002). Measurement of ATP production and respiratory chain enzyme activities in mitochondria isolated from small muscle biopsy samples. *Anal Biochem* 311, 139-151.

Wilbe, M., Jokinen, P., Truve, K., Seppala, E.H., Karlsson, E.K., Biagi, T., Hughes, A., Bannasch, D., Andersson, G., Hansson-Hamlin, H., Lohi, H. & Lindblad-Toh, K. (2010). Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. *Nat Genet* 42(3), 250-4.

Acknowledgements

Göran Andersson, main supervisor. Thank you for encouraging my scientific development. I appreciate that you allowed me to choose my own path and I am grateful for the trust and freedom you placed in me.

Leif Andersson, co-supervisor. Your passion for science is inspiring and the width of your knowledge is impressive. Thank you for always taking time to discuss my projects; you make me want to know more and work harder.

Kerstin Lindblad-Toh, co-supervisor. You are the hardest working person I have, and probably will ever, meet. I'm impressed with your devotion to help mankind! Thank you for introducing me to the Broad community.

Gerli Rosengren Pielberg, co-supervisor. What would this lab be without you? You are one of the smartest and most pedagogic people I know. You are my dear friend and I appreciate your honesty!

Åke Hedhammar, co-supervisor. "The spider in the web", everyone in the "dog-world" knows you :-) Thank you for sharing your very valuable contacts.

Karin Hultin Jäderlund, co-author. I could have never asked for a better collaborator and friend. Your knowledge in veterinary neurology is impressive, as is your desire to solve the different aspects of SAN and other neurological disorders. I would love to work with you on other projects!

Rolf Wibom, co-author. Thank you for introducing me to the mitochondrial world in humans.

Elinor K. Karlsson, thank you for generously sharing your computational and bioinformatics knowledge. You make complicated things seem easy.

Many thanks to all other co-authors involved in the MITF and SAN-stories.

Jennifer Meadows, thank you for spending many hours reading this thesis and providing many valuable comments, both scientific and linguistic. I enjoy our times spent in the lab, I love your sarcastic sense of humour.

My roomies, past and present. Greger Larsson, I miss your presence in the lab. The girls: Mia Olsson, Nina Sundberg and Katarina Tengvall, I could not have asked for better roomies and friends.

Ulla Gustafson, you are one of the solid points of this lab. Thank you for always finding time to give me a hand in the lab when things get stressful.

Gudrun Wieslander and Helena Pettersson, you are the best administrators one could ask for!

Patricio Rivera, Tove Fall and all other veterinary colleagues at the animal hospital, Ultuna. Thanks for sharing your knowledge and good times.

Past and present members of the lab, none mentioned and none forgotten.

All dog owners and breeders who kindly provided samples included in these studies.

Per Jensen, you introduced me to the world of science and research, and you gave me the confidence to find my own path. Thank you!

Kjerstin Pettersson, you are the kindest and most empathetic person I know. You were there when I needed you, if you ever need anything...

My parents, pappa, mamma och Mats, thank you for always being there for me, I can always trust and rely on you no matter what!

Min familj, Alicia och Markus, vad skulle jag göra utan er? Tack för ert gränslösa tålamod med en stundtals stressad och upptagen mamma och fru. Jag älskar er!