Genetic Studies in Dogs Implicate Novel Genes Involved in Atopic Dermatitis and IgA Deficiency

KATARINA TENGVALL
Dissertation presented at Uppsala University to be publicly examined in B22, BMC, Husargatan 3, Uppsala, Tuesday, 6 October 2015 at 13:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Docent Ingrid Skelton Kockum (Clinical Neuroscience (CNS), Karolinska Institute).

Abstract

This thesis presents genetic studies of atopic dermatitis (AD) and IgA deficiency in dogs.

AD is a chronic inflammatory and pruritic skin disorder caused by allergic reactions against environmental allergens. Both genetic and environmental factors are involved in the development of Canine AD (CAD) and human AD. In Paper I, we performed genome-wide association studies (GWAS) and identified a locus on chromosome 27 significantly associated with CAD in German shepherd dogs (GSDs). The locus contains several genes and fine-mapping indicated strongest association close to the candidate gene PKP2. In Paper II, we performed additional fine-mapping and identified four highly associated SNPs located in regions with transcriptional regulatory potential in epithelial and immune cells. The risk alleles were associated with increased transcriptional activity and the effect on expression was cell-type dependent. These data indicate that multiple cell-type specific enhancers regulate the expression of PKP2, and/or the neighboring genes YARS2, DNM1L and FGD4, and predispose GSDs to CAD.

IgA deficiency is the most common primary immune deficiency disorder in both humans and dogs, characterized by a higher risk of recurrent mucosal tract infections, allergic and other immune-mediated diseases. In Paper III, we performed the widest screening (to date) of serum IgA levels in dog breeds (N_dogs=1267, N_breeds=22) and defined eight breeds as predisposed to low IgA levels. In Paper IV, we performed GWAS in four of the breeds defined as prone to low IgA levels. We used a novel percentile groups-approach to establish breed-specific cut-offs to perform analyses in a close to continuous manner. In total, 35 genomic loci were suggestively associated (p<0.0005) to IgA levels, and three genomic regions (including the genes KIRREL3 and SERPINA9) were genome-wide significantly associated with IgA levels in GSDs. A ~20kb long haplotype on chromosome 28, significantly associated to IgA levels in Shar-Pei dogs, was positioned within the first intron of the gene SLIT1 overlapping with a possible dog domestication sweep.

This thesis suggests novel candidate genes involved in two immune-mediated disorders in the dog. Hopefully, these results will become an important resource for the genetic research of the corresponding human diseases.

Keywords: GWAS, canine model, genetic association, immunogenetics, atopic dermatitis, IgA deficiency

Katarina Tengvall, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.

© Katarina Tengvall 2015

ISSN 1651-6206
urn:nbn:se:uu:diva-259606 (http://urn.kb.se/resolve?urn=nbn:se:uu:diva-259606)
Till Pappa
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


II Tengvall, K., Kozyrev, S.*, Kierczak, M.*, Bergvall, K.#, Farias, F.#, Ardesjö-Lundgren, B., Murén, E., Hagman, R., Leeb, T., Pielberg, G., Andersson, G., Hedhammar, Å., Lindblad-Toh, K. A risk haplotype within the PKP2 locus shows association to Canine Atopic Dermatitis and contains cell-type specific enhancers. (Manuscript) *These authors contributed equally to this work. #These authors contributed equally to this work.


IV Tengvall, K.*, Olsson, M.*, Frankowiack, M.#, Kierczak, M.#, Bergvall, K., Axelsson, E., Tintle, L., Marti, E., Roosje, P., Leeb, T., Hedhammar, Å., Hammarström, L., Lindblad-Toh, K. Genome-wide Analyses Suggest Mechanisms Involving Early B-cell Development in Canine IgA Deficiency. PLoS One, 10(7):e0133844. doi: 10.1371/journal.pone.0133844 *These authors contributed equally to this work. #These authors contributed equally to this work.

Reprints were made with permission from the respective publishers.
Contents

Introduction .................................................................................................................. 13
  Genetic studies of the dog ......................................................................................... 14
    The dog genome ........................................................................................................ 15
  Genome-wide association studies ........................................................................... 17
    The dog as a genetic model for human disease .................................................. 21
  Genetic mapping of complex traits ......................................................................... 23
Introduction to immunology ....................................................................................... 25
  The immune response ............................................................................................... 25
  Lymphocytes ............................................................................................................. 27
  Hypersensitivity reactions ......................................................................................... 29
  Primary immunodeficiencies ..................................................................................... 30
Atopic dermatitis ......................................................................................................... 30
  The immune response underlying atopic dermatitis ............................................. 31
  Canine atopic dermatitis ......................................................................................... 32
  Comparisons between canine and human atopic dermatitis .................................. 34
  Environmental factors ............................................................................................. 37
  Genetics .................................................................................................................... 37
Immunoglobulin A deficiency ...................................................................................... 39
  Human IgA deficiency ............................................................................................... 40
  Low IgA levels in dogs ............................................................................................. 41
Aims of the thesis .......................................................................................................... 42

Present Investigations .................................................................................................. 43
  Papers I & II: CAD-associated genetic variants are located in tissue-specific enhancers within the PKP2-locus and participate in transcriptional up-regulation ................................................................. 43
    Background ............................................................................................................. 43
    Paper I: Genome-wide association of a locus on chromosome 27 with CAD in German shepherds ................................................................................................................. 44
    Paper II: Four CAD-associated SNPs within the PKP2-locus show regulatory effects on gene transcription ................................................................................................. 46
  Discussion ................................................................................................................ 49
Papers III & IV: Genome-wide analyses suggest novel genes involved in the regulation of serum IgA levels in German shepherds and Shar-Pei

Background ................................................................................................................. 52

Paper III: Serum IgA screening of multiple dog breeds identifies breeds prone to low IgA levels ................................................................................................. 53

Paper IV: IgA-associated loci harbor candidate genes with indicated functions in early B-cell development ................................................................. 55

Discussion ...................................................................................................................... 57

General discussion ........................................................................................................ 60

CAD candidate genes ............................................................................................... 60

IgA candidate genes ................................................................................................. 62

IgA and CAD ............................................................................................................. 64

Connection between CAD and IgA ........................................................................... 64

Clinical presentation of CAD in different breeds .................................................. 65

IgA in wolves ............................................................................................................ 65

Future perspectives .................................................................................................... 66

Populärvetenskaplig sammanfattning ..................................................................... 68

Bakgrund ............................................................................................................... 68

Nya forskningsresultat ............................................................................................. 69

Acknowledgements ..................................................................................................... 72

References .................................................................................................................. 74
Related Work by the Author

(Not included in the thesis)


## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>Aplastic anemia</td>
</tr>
<tr>
<td>AD</td>
<td>Atopic dermatitis</td>
</tr>
<tr>
<td>APC</td>
<td>Antigen presenting cell</td>
</tr>
<tr>
<td>ARVC</td>
<td>Arrhythmogenic right ventricular cardiomyopathy</td>
</tr>
<tr>
<td>ASIT</td>
<td>Allergen specific immunotherapy</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair(s)</td>
</tr>
<tr>
<td>CAD</td>
<td>Canine atopic dermatitis</td>
</tr>
<tr>
<td>CAFR</td>
<td>Cutaneous adverse food reaction</td>
</tr>
<tr>
<td>CNV</td>
<td>Copy number variant</td>
</tr>
<tr>
<td>CVID</td>
<td>Common variable immunodeficiency</td>
</tr>
<tr>
<td>Gb</td>
<td>Billion base pairs</td>
</tr>
<tr>
<td>GSD</td>
<td>German shepherd dogs</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association study(ies)</td>
</tr>
<tr>
<td>HRCAD breed</td>
<td>High risk CAD breed</td>
</tr>
<tr>
<td>IDECs</td>
<td>Inflammatory dendritic epidermal cells</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IF</td>
<td>Intermediate filament</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase pairs</td>
</tr>
<tr>
<td>LCs</td>
<td>Langerhans cells</td>
</tr>
<tr>
<td>LRCAD breed</td>
<td>Low risk CAD breed</td>
</tr>
<tr>
<td>MAF</td>
<td>Minor allele frequency</td>
</tr>
<tr>
<td>Mb</td>
<td>Megabase pairs</td>
</tr>
<tr>
<td>MDS</td>
<td>Multidimensional scaling plot</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>PAMPs</td>
<td>Pathogen-associated molecular patterns</td>
</tr>
<tr>
<td>PID</td>
<td>Primary immunodeficiency</td>
</tr>
<tr>
<td>SCID</td>
<td>Severe combined immunodeficiency</td>
</tr>
<tr>
<td>SLE</td>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>TGF</td>
<td>Transforming growth factor</td>
</tr>
<tr>
<td>Th cell</td>
<td>T helper cell</td>
</tr>
<tr>
<td>TSLP</td>
<td>Thymic stromal lymphopoietin</td>
</tr>
</tbody>
</table>
Introduction

Biomedical research has shown a remarkable development over the last decade, especially when it comes to the mapping of genetic variants underlying traits or diseases. A significant milestone of this new era came in 2001 when the first draft of the human genome sequence was published [1, 2] after more than a decade of work at a cost of around 3 billion USD [2]. The human genome project yielded important new knowledge about the architecture and function of the genome including protein-coding gene number and density, non-coding genes, repeat sequences, and evolutionary conservation [3]. Today, a human genome sequence can be delivered at a cost of around $10,000 [4] in just one day. Indeed, the technical development behind whole-genome sequencing has been extraordinary. Shortly after the human genome sequence release, the mouse [5], rat [6], and dog [7] genomes were published. When comparing these four genomes it was revealed that the gene count is very similar across the species (~20,000) and at least 5% of the genome is under selection. Only 1.5% of the genome consists of protein-coding genes, whereas the rest of the sequences under selection likely harbor important regulatory elements. In 2011, a comparative analysis of sequenced genomes from 29 (placental) mammals revealed that around 4.2% of the genome contains highly conserved genomic elements, most likely with functional properties owing to their conservation across species. Potential functions were assigned for more than half of these elements [8]. Currently, the genomes of 50 mammals have been sequenced and the aim is to finish sequencing an additional 150 mammals within the coming year, using newly developed technology that is faster and less expensive.

Thanks to the continuous update of genome sequences from species both closely and distantly related to human, we constantly learn more about genome evolution and what is in the human genetic code. Today, the problem is not the lack of genetic information, instead the biggest challenge is to handle enormous amount of genomic data to interpret it and understand the functions of conserved elements. The advances in genotyping technologies and the completion of genome assemblies have been important prerequisites for the use of genome-wide association studies (GWAS) to map genes underlying different traits, such as diseases. The ultimate goal of current genetic research is to explain the biological pathways underlying certain traits.
and, in particular, various diseases to find ways to prevent, diagnose, treat and cure them.

In this thesis, I will present the work from genetic studies of two immune-mediated diseases: **atopic dermatitis** – when the immune system becomes overly sensitive and reacts to harmless environmental particles, and **IgA deficiency** – when the immune system has a shortage of antibodies that are important for identifying and neutralizing infectious agents. Both diseases show high clinical similarities between humans and dogs, and the genetic background is still largely unresolved. The studies were conducted in the dog with the intention that the results will become a resource for the genetic research of the corresponding diseases in human. However, the primary goal of this thesis was to identify novel genes involved in atopic dermatitis and IgA deficiency in the dog. This new information will hopefully assist in the development of disease treatments and diagnostics, and also in breeding for improved canine health. GWAS was the key method of choice, resulting in the successful mapping of associated genomic regions harboring candidate genes including **PKP2** (atopic dermatitis in the German shepherd dog), **KIR-REL3** (IgA deficiency in the German shepherd dog), and **SLIT1** (IgA deficiency in the Shar-Pei). To reach this goal there were many steps on the way, including phenotypic classifications, sample collection and preparation, data handling, statistical modeling, pathway analyses as well as detection of causative variants. Within this introduction I will give brief overviews covering various disciplines involving medicine, genetics, biotechnology, statistics, and immunology, to provide a comprehensive background summary to the four studies that build this thesis.

### Genetic studies of the dog

The domestic dog (*Canis familiaris*) has been humans’ prime accompanying animal for thousands of years and has been used for various purposes including hunting, herding, guarding, and protection. The dog was domesticated from the grey wolf (*Canis lupus*) at least 15,000 years ago at multiple times and locations in East Asia, however the exact times and places for dog domestication are still debated [9]. The creation of the majority of modern dog breeds started more recently, around 200 years ago [10]. By keeping registries and issuing pedigrees, the controlled breeding in closed populations has resulted in hundreds of different breeds. The intense artificial selection (when humans select individuals carrying a certain phenotype, such as a behavioral characteristic, for breeding) became highly efficient in dog breeds through backcrossing (offspring-parent mating), few founders (when only a handful of dogs were used to initiate an entire breed), and the use of matadors (few intensely used stud dogs). This has created the wide morpho-
logical and behavioral variation between breeds that is present today. The modern dog is actually the most morphologically diverse animal of the domesticated species [11], and there are around 400 different breeds recorded today, in addition to several mixed breeds and feral dogs [12]. A consequence of this controlled breeding is the accelerated pace of fixation of genetic variants and genetic drift (random loss of genetic variants). All individuals normally carry rare deleterious genetic variants, which in wild populations will be eliminated through natural (purifying) selection. Inbreeding and artificial selection may instead lead to the enrichment of disease-causing mutations either directly (pleiotrophy) or indirectly (hitch-hiking). Pleiotrophy occurs when the genes responsible for the desired phenotypes also have other effects, such as causing disease, through separate biological mechanisms [13], whereas hitch-hiking is caused by linkage on the chromosomes when disease-causing mutations are inherited together with the genetic variants that cause the desired phenotypes [14]. Many of the dog breeds show high susceptibility to one or more diseases, which typically differ between breeds. For example the Irish wolfhound, Cavalier King Charles spaniel, and the Great Dane display a 15, 11 and 9-fold risk, respectively, to develop cardiovascular disease compared to the average dog breed population [15], whereas the Australian terrier and Samoyed show a 10 and 7-fold risk, respectively, to develop diabetes mellitus [16]. The high risks indicate strong genetic risk factors underlying these diseases in dogs, which will thereby improve the chances that the genetic alterations will be identified in genetic studies.

The dog genome

There has been a pronounced breakthrough of genetic mapping of traits in the dog since the complete and high quality genome sequence of the female Boxer “Tasha” was published in 2005. The dog genome sequence is about 2.4 billion bases (Gb) divided up on 38 autosomal chromosomes and chromosome X. A little more than 2.5 million single nucleotide polymorphisms (SNPs) were identified in the dog genome through comparisons to low coverage draft genome sequences of other breeds and canids, as well as by comparing the chromosome pairs of the Boxer genome sequence [7]. SNPs are single bases that vary between individuals (i.e. are polymorphic) and typically show a frequency of the least common allele i.e. minor allele frequency (MAF) of more than 1% within a population. SNPs represent one type of mutation, which refers to any change in the DNA sequence that deviates from the set reference sequence Thus, a mutation is the rare and abnormal variant in contrast to the definition of a SNP, which is more common in a population. SNPs are nowadays the most commonly used genetic markers in GWAS. The first SNP array developed for genotyping dogs consisted
of ~27,000 SNPs [17]. Since then, the genome sequence of the dog has been improved and the marker map has increased to contain around 170,000 SNPs, \textit{i.e.} the canine HD Illumina 170K SNP array, in order to ensure sufficient coverage across the various dog breeds [18].

The genome of the domestic dog is shaped by two population bottleneck events (Figure 1) resulting in a mosaic pattern in regards to the haplotypes. Haplotypes are stretches of DNA segments that seldom get separated by \textit{recombination}, which is when new chromosomal combinations are formed during meiosis by crossing-over events in the gametes and then passed on to the offspring. There are recombination hotspots (with high recombination rates) in the genome and some regions, between the hotspots, remain the same after many generations of recombination. These regions are called \textit{haplotype blocks} and are the signature pattern of large and old populations.

The \textit{first bottleneck} occurred at domestication when the domesticated dog species was created from a subset of wolves. Only a part of the gene pool of the large wolf population (harboring short haplotype blocks) became the genome of the dog, hence the bottleneck. The \textit{second bottleneck} happened when the modern dog breeds were created and long haplotypes became typical features within breeds. Due to the short time since breed creation, the long haplotypes have not yet been substantially broken down by recombination, and since inbred populations carry long stretches of homozygosity, recombination events in these regions are hidden. The ancestral short haplotype blocks are shared across breeds. This mosaic pattern becomes particularly clear when taking the study of white coat color in Boxers and Bull terriers as an example. Here the haplotype defining white coat color in Boxers was about 1Mb long, whereas the haplotype harboring the genetic variant causing the white color, shared across the two breeds, was instead around 100kb long [17].
Figure 1. Mosaic pattern of the dog genome formed by two genetic bottleneck events. The dog genome is shaped by two genetic bottleneck events. (a) The pre-breed dogs carried short haplotype blocks derived from the wolf population at the time of domestication. (b) The selection for certain characteristics started more recently and the different breeds that were created shared long haplotypes. (c) Taken together, breeds have long haplotypes but the short ancestral haplotypes are shared across breeds in a mosaic looking pattern (Modified from Karlsson & Lindblad-Toh 2008) [10].

Genome-wide association studies

In GWAS, each single marker is tested for the association to the phenotype, *e.g.* a disease by using a group of cases affected by the disease and compare to a group of healthy controls. The associated markers pinpoint the genomic region(s) where cases and controls are different. The assumption made when performing GWAS is that the causative mutation is *identical-by-descent* within a population (*e.g.* a dog breed). The markers used in GWAS are usually SNPs and to associate a *locus* (*i.e.* genomic region) to the phenotype a sufficiently dense SNP map is required. SNPs on the same haplotypes are in *linkage disequilibrium* (LD) with each other, meaning that certain *alleles*...
(i.e. genetic variants) at each SNP position are inherited together more often than by random chance. The SNP map requires SNPs that are tagging all haplotypes in order to cover and represent the genome sufficiently. In dog GWAS, calculations have estimated that 73% of the SNPs are polymorphic (MAF>5%) within a typical breed, therefore the SNP-chip is designed to include an excess of SNPs to account for the loss of the monomorphic SNPs [7]. In GWAS, typically the additive effect (to some extent ignoring the dominance effect) of each single SNP is taken into account, by using the following (simplest) model:

\[ y = X\beta + e \]

where \( y \) is the phenotype vector and \( X \) is a matrix where each row represents one individual and each column a fixed effect like the genotype (e.g. coded 0, 1 and 2) of the tested SNP and additional parameters influencing the trait for example sex and age. The \( \beta \) vector (including the effect of the population mean, \( \beta_0 \)) represents the additive SNP effect and the additional fixed effects. The residual \( e \) (error) is simply the difference between the actual (observed) phenotype \( y \) and the phenotype predicted by the model (\( \hat{y} \)), i.e. what the model cannot explain. This residual is due to the factors not included in the model, e.g. polygenic effects, unknown environmental factors or interactions of these variables. The smaller the \( e \) the more of the phenotypic variance is explained by the model (including the tested SNP). However, even if \( e \) is large, the SNP may still be significantly associated to the phenotype if there is, for example, a high environmental influence that is not included in the model.

**Significance thresholds**

Next, by performing a statistical test (e.g. \( \chi^2 \) (Chi-square) in case versus control studies, or Student’s t-test or Armitage trend test for continuous traits) on \( \beta \), a p-value for each SNP is obtained. Normally the threshold of significance is set to \( p=0.05 \) (the nominal 5% threshold), which means that in 5% of the events the null hypothesis is falsely rejected i.e. resulting in a false positive result. In GWAS, there is a challenge caused by multiple testing since the risk of detecting an association due to chance increases with the number of independent tests (in this case the number of SNPs). To account for this, the Bonferroni correction [19] is traditionally used and defines a new and stricter significance threshold by dividing 0.05 by the number of tests. However, if taking LD into account each SNP is actually not an independent test. Therefore, the Bonferroni approach often results in an overcorrection in GWAS. One way to work around this is to estimate the number of independent tests based on not only the number of SNPs, but also the LD pattern. This was done for dog GWAS by Karlsson *et al.* 2013, where the threshold for suggestively associated SNPs was set to \( p=5\times10^{-4} \) based on the
assumption that haplotype blocks in a dog breed are $\sim 1\text{Mb}$ long [20]. For comparison, a threshold for suggestive association of $p = 10^{-5}$ has been used in human [13]. Another preferable alternative to find genome-wide significant associations is to use permutations, where the association analysis is rerun multiple times (usually 1,000 up to 1,000,000) with a random reshuffling of the phenotypes within the sample cohort. Such an approach determines the threshold where there will be a 5% risk (set at 5% by tradition) or less to have a false positive result, determining the genome-wide significance threshold specific for the current dataset. Any SNP association with p-values below this threshold, referred to as the empirical threshold, will be considered genome-wide significant.

Covariates and population stratification

Sometimes, one of more known factors (fixed effects) may influence the trait. Known fixed effects can be taken into account in the statistical model by introducing them as covariates. Covariate adjustment reduces spurious associations due to fixed effects but it may also reduce the statistical power. Fixed effects can be sex, age, clinical variables or population stratification. The latter situation being when the individuals within the sample set are unevenly related across the groups of cases and controls, resulting in false associations, i.e. associations unconnected to the phenotype of interest. Assuming no stratification, the observed p-values should perfectly follow the expected p-value distribution (except for the signal(s) of association). In case of population structure and spurious association signals, many of the p-values in the observed distribution will deviate from the expected. The comparisons of distributions are usually presented as quantile-quantile (QQ) plots and the deviation is typically measured by the genomic inflation factor $\lambda$, which is the regression slope (angle) of the line fitted to observed p-values compared to the null (i.e. expected) [21]. The simplest solution to correct for genomic inflation is to use genomic control, which simply restores the expected distribution by dividing the observed p-values by the $\lambda$ [22]. To account for population stratification as well as additional covariates, a mixed model can be used:

$$\gamma = X\beta + Zu + e$$

where $X$, $\beta$ and $e$ are the same as in the previous model. The $Z$ is the genomic kinship matrix, which captures the genetic relationship between individuals based on the genotype data, and the $u$ is the distribution of values (effects) that defines how this genetic similarity contributes to the phenotype. Instead of adding a vector of effects (one effect per covariate) as in $\beta$, the $u$ is the distribution of effects influencing the trait, which is much more computationally efficient. The $u$ is assumed to follow a normal distribution characterized by its mean and variance. This also adds more flexibility into
the model as it enables the analyses of variables where not all the values are known/observed. For instance, if the effect of three different diets are added but it is known that there may be additional diets not yet observed this can be modeled in the random part of the model (Zu) instead of in the fixed part (Xβ). Also, instead of accounting for the effect from other SNPs one by one (inefficient in a GWAS with a large number of SNPs), the polygenic effect of all SNPs in u can be captured by setting Z to be the genomic kinship matrix. In summary, the mixed model efficiently corrects for cryptic relationships (typical to dog breeds) and a certain degree of population stratification through the use of the genomic kinship matrix.

Subpopulations
In addition to mild population stratification, there can also be an issue with the formation of subpopulations. This is typically seen between dog breeds but may also distinguish different types within breeds or geographically separated populations. The genomic kinship matrix may not capture this and will only correct for cryptic relatedness within each cluster. In this case an extra variable needs to be added to X with the cluster information. This advances the mixed model into a mixed model for stratified association. To judge if there are subpopulations or not, one can plot the genomic kinship matrix in a multidimensional scaling (MDS) plot. This type of plot contains one point per individual and the coordinates are assigned to each individual in such a way that the distance between the points reflects the genetic distance/similarity between the individuals. It is then possible to assign clusters visually or by using more advanced statistical methods (e.g. the scree test).

GWAS in human versus dog
In human, the initial International HapMap projects aimed at defining LD-patterns by genotyping about three million SNPs in four human populations [23, 24]. It was estimated that approximately half a million SNPs were needed to adequately capture the human DNA sequence in GWAS [24]. A series of SNP microarrays have been developed for that purpose and almost 2,000 GWAS studies have been published since 2005 (http://www.genome.gov/gwastudies/), primarily using Illumina or Affymetrix SNP genotyping arrays. The newer Illumina HumanOmni2.5 and Omni5.0 (2.5 and 5.0 million SNPs, respectively) were launched in 2010 with an expanded coverage to enable inclusion of less common SNPs (MAF of 1-5%) [3]. Since the dog breeds share longer haplotypes compared to the human population, dog GWAS can be performed using fewer SNPs to capture all haplotypes of the genome. The current SNP-chip developed for dogs include 170K SNPs compared to 2.5-5M SNPs on the human SNP arrays.
The statistical power obtained in a GWAS is influenced by multiple factors: sample size, causal allele frequency, marker allele frequency, the linkage between the causal mutation and marker, and also the effect size. The latter is basically the Odds Ratio (OR) measured as the fold risk for an individual carrying the specific risk factor to develop a disease/trait compared to an individual without the risk factor. Most often, several loci influence complex traits in humans and the majority of the effect sizes are small (OR ≤1.5), and even modest effect sizes (OR of 1.5-2) require large sample sizes [13]. The corresponding complex diseases in dogs are typically conferred by a higher risk within a dog breed. Power calculations in the dog suggest ~100-300 cases and ~100-300 controls for GWAS of a complex disease, with a 5-fold increased risk of developing disease for individuals carrying the genetic risk factors [17]. Around 100 times more study subjects are required to reach the sufficient power in an association study in human [13] compared to dogs, due to the longer haplotypes and larger effect sizes in dogs [17]. Thus, both the estimated number of samples and SNPs required to detect significant associations are lower in dog GWAS compared to human GWAS.

The dog as a genetic model for human disease

The dog is similar to human in many ways and the resemblances between canine and human diseases are striking [25]. Actually, there are around 360 genetic diseases described in human where the dog may be a potential model (http://omia.angis.org.au/home/) [26] due to the clinical resemblance. To establish a good animal model for mapping diseases and traits that will benefit human genetic research, many aspects need to be considered.

Comparisons to the mouse model

The genomes of dog and human are more similar to each other compared with human versus mouse or mouse versus dog. Despite the fact that rodents split from the ancestral mammalian branch later than the split between the carnivore and primate lineages, the mouse genome is more diverged from human compared to dog. This is mainly because of the shorter generation time resulting in a faster evolution rate in rodents. Dogs have basically the same number of genes as humans, with the majority being 1:1 orthologues (i.e. descended from the same ancestral gene) [10]. Genetic diseases in dogs are spontaneously occurring, in contrast to the laboratory mouse strains where the studied diseases are induced chemically or achieved from directed selection to create inbred mice strains where all individuals within the strain are affected by the disease. The genetic mouse model provides efficient testing of the effect of major genes. However, the mouse model is limiting when searching for novel gene associations of complex human diseases, which most often have a polygenic nature. Dogs on the other hand, despite being
artificially selected for favorable traits, provide a natural disease model for both multi- and monogenic diseases. The registered pure-breed dogs also have pedigrees enabling researchers to track the family history of different diseases providing valuable information regarding the mode of inheritance of the disease of interest. Moreover, dogs routinely receive medical care for many of the common diseases in a manner similar to human. This provides the possibility to record symptoms and the effects of treatment [10]. In addition, the importance of the environmental influence on the development and progress of a complex disease must not be forgotten. Whereas mice used for research are usually kept in highly controlled environments, the dogs live right among us sharing our environment including food, living environment, habits of physical activity, and exposure to factors like tobacco smoke.

Despite the drawbacks of the induced mouse model to study complex diseases as mentioned above, it has a long history of significant contribution to biomedical research. Advantages of the mouse as a model include the possibility of genetic manipulation (such as knockouts where genes are either completely or partially disabled), the short generation time, controlled crosses to evaluate the effect of mutations in different genetic backgrounds, low housing costs, and controlled environments. The inbred mouse strain NC/Nga is one example of the usefulness of a mouse model for a complex disease. The NC/Nga mouse was studied as a model for human atopic dermatitis (AD) already 15 years ago as it spontaneously developed skin lesions when maintained under conventional but not pathogen-free environment [27]. The NC/Nga mouse model is still used extensively in studies of AD, for example when evaluating treatments [28]. Another AD mouse model is the Ovalbumin induced model (*i.e.* the epicutaneous sensitization model), which mimics skin lesions similar to those observed in human AD [29].

Disadvantages with the dog as a model organism primarily involve complex relatedness and sample collection. Despite focusing on collecting samples from unrelated dogs, the history of breed creations becomes apparent as cryptic relatedness. This is common among dogs within a breed and may complicate the data analyses. The mixed model can help to a great extent, but finding true associations can sometimes be problematic. Moreover, despite the need for fewer samples compared to human studies, it may be difficult to collect sufficient numbers of cases and controls in small populations as well as finding healthy controls in a breed with a high prevalence of the disease of interest. Finally, once an association is detected, the extended haplotypes within breeds may be difficult to fine-map unless there is another breed available with the same phenotype and where the causative genetic variant(s) is(are) identical-by-descent.
Genetic mapping of complex traits

The mapping of complex diseases, which are caused by multiple genes/mutations and most often influenced by environmental factors, is perhaps the biggest challenge in medical genetics. Mutations can be SNPs, insertions or deletions of single nucleotides or larger DNA segments, and duplications resulting in copy number variants (CNVs). When mutations occur in functionally important genetic regions, either coding or cis-regulatory elements, they can result in phenotypic changes (Figure 2). Cis-regulatory elements are DNA-stretches, including promoters, enhancers, and silencers that regulate the transcription of genes. Mutations that cause simple Mendelian traits are usually protein-coding mutations, resulting in a dysfunctional protein across all tissues where it is expressed. In complex diseases, the majority of the mutations are in regulatory elements with an effect limited to certain tissues or developmental stages depending on the location of the mutation. These mutations may alter the binding capability of RNA-polymerase and transcription factors, thus affecting the expression of the gene.

Figure 2. Mutations located in the coding regions and cis-regulatory elements.

Certain DNA segments around and within genes function as regulatory elements (including enhancers, silencers and promoters) as transcription factors specifically bind to their DNA sequences and together control gene expression. Mutations (*) resulting in phenotypic changes can be located either in coding regions or in regulatory elements. This figure is kindly provided by Dr J. Eriksson and modified from Coyne and Koekstra (2007) [30].
In human, the small effect sizes of most associated SNPs (OR<1.5) in GWAS account for only a small proportion of the heritability (i.e. the proportion of the phenotype explained by genetics) of the studied complex traits. This limits the applicability of the identified mutations for risk predictions. Furthermore, a large proportion of the genetic risk is still unexplained, often referred to as the missing heritability. The missing heritability has been widely discussed and there are various strategies proposed to improve the methodology. These include: increasing sample sizes by combining GWAS (meta-analyses), investigating rare variants through targeted/exome sequencing, and evolving the more complicated methodologies that take into account epistasis and epigenetic effects on complex traits [3]. Two opposing models have been proposed to explain the missing heritability: the common disease/common variant (CD/CV model) and the common disease/rare variant (CD/RV model) hypotheses. The proportions between the two are likely to vary between different complex phenotypes. Detection of less-common SNPs and CNVs involved in diseases has just started to accumulate in the scientific literature. The increased coverage of SNPs on the human SNP-chips will also aid in detecting more common variants than previously. However, the high density SNP-chips also require even larger sample sizes to detect the associations.

When a study’s result is replicated in another population it provides a strong indication of a true finding. However, if a result is not replicated, this does not prove that it is false. Instead it might solely reflect the differences between populations when it comes to genetic factors influencing a trait. In dogs, the mapping of the white coat color locus shared by Boxers and Bullterrier was one of the first proof-of-principle studies suggesting simple across-breed fine-mapping to identify the major disease-causing mutations identical-by-descent between breeds. However, subsequent dog studies have experienced a more complex pattern. Despite multiple breeds being predisposed to the same disease, the genetic risk factors are not necessary shared. For instance, GWAS of three breeds predisposed to osteosarcoma identified 33 risk loci, whereas none was shared across the breeds but instead harbored genes in the same pathways [20]. Canine leukocyte adhesion deficiency (CLAD) in Irish setter [31], Shar-Pei fever [32], and systemic lupus erythematosus (SLE)-like disease in the Nova Scotia duck tolling retriever breed [33] are examples of diseases so far known to be limited to one breed only. Also, breeds may be fixed for causative mutations and lack informative segregating SNPs resulting in no association signal within that breed. Such a homozygous region, or selective sweep, may instead be detected when searching for regions of reduced heterozygosity. The Shar-Pei study provides a typical example where the association signal was partly hidden by a
selective sweep due to the strong selection for the thick and wrinkled phenotype, characteristic for the Shar-Pei [32].

To summarize, the finding of a genetic association to a disease in the dog does not prove that the same genes are also involved in the corresponding disease in human. However, there is a strong possibility that the genes found in the dog will increase the knowledge about the biological pathways underlying certain diseases and/or be directly involved in subsets of heterogenic human populations. Indeed, this will contribute to the research of human medical genetics.

Introduction to immunology

The body is constantly exposed to pathogens, defined as infectious agents (e.g. viruses, bacteria, parasites, or fungi) that can give rise to disease. Fortunately, the immune system has emerged in the course of evolution and provides a remarkable protection against infectious agents to prevent disease. The immune system can be divided into the innate and adaptive immune responses, which primarily differ in terms of time and memory. This section aims at briefly describing the immune response to provide an insight into the extraordinary complexity of our immune system, as a background for the research presented in this thesis.

The immune response

The principal cells of the immune system are the white blood cells that are called leukocytes. Leukocytes are constantly produced by the hematopoietic stem cells in the bone marrow. These multi-potential hematopoietic stem cells give rise to the myeloid (monocytes/macrophages, neutrophils, basophils, eosinophils, megakaryocytes/platelets, and dendritic cells) and the lymphoid (T and B lymphocytes and natural killer cells) progenitor cell lineages. Cytokines, including chemokines, interleukins (ILs), and interferons (IFNs), are small proteins secreted by cells and affect the behavior of nearby cells that carry the appropriate receptors. Cytokines are essential for immune cell signaling and communication, and induce for instance growth, differentiation, or death of the target cells [34].

The innate immune system

The epithelial surfaces (such as the skin) and the mucosa are physical barriers that are highly efficient as the first line of defense against the intruding pathogens. However, if pathogens do invade the body, the innate immune system reacts fast and unspecific initiating inflammation. This is typically
started by immune and/epithelial cells that recognize pathogen-associated molecular patterns (PAMPs), which are molecules generally shared by pathogens. Inflammation is characterized by the local accumulation of fluid, plasma proteins, and leukocytes and the typical clinical signs are pain, redness, and swelling. The cells of the innate immune system are primarily phagocytic immune cells, which basically engulf the pathogen in a process called phagocytosis. The main cells of the initial phase of the inflammatory response are the numerous neutrophils followed by macrophages, which release inflammatory cytokines as well as cytokines that attract other immune cells from the blood stream to the site of infection. The invading pathogens also trigger plasma proteins to activate the complement system. The complement system initiates inflammation through a cascade of proteolytic reactions ending up with a specific protein (C3b) that attaches to the pathogen surface assisting phagocytosis [34].

The adaptive immune system

Immature dendritic cells act as antigen presenting cells (APC) when they phagocytose the pathogens at the site of infection and migrate to the peripheral lymphoid organs to present antigens to naïve T lymphocytes. Antigens (abbreviation of “antibody generator”) are molecules that can be recognized by the highly variable antigen receptors on T or B lymphocytes. The APCs present specific peptide antigens on the major histocompatibility complex (MHC) class I and class II molecules. The MHC molecules are highly polymorphic glycoproteins expressed on the cell surface and are specialized in binding peptide fragments from pathogens. The MHC class I molecules are expressed on basically all nucleated cells except neurons, whereas the MHC class II molecules are more specific and expressed primarily on specialized APCs such as dendritic cells, macrophages, and B lymphocytes. The activation of T lymphocytes initiates the adaptive immune response. This is the delayed and more specific response with many of the same effector mechanisms as the innate immune system, but with the ability to target specifically infected cells through a: i) MHC class I-dependent mechanism whereby viral peptides from the infected cells are presented by MHC class I molecules to cytotoxic CD8+ T killer cells, and ii) MHC class II-dependent antigen presentation to CD4+ T helper cells from specialized APCs of antigens derived from exogenous bacteria, viruses or parasites. Both these antigen presentations occur with great precision because of affinities between specific antigens and the specific alleles of MHC class I and class II expressed by the individual [34].
Lymphocytes

This section will focus on the main players of the adaptive immunity: the different types of T lymphocytes (T-cells) and the antibody producing B lymphocytes (B-cells). These express a huge repertoire of antigen receptors: T cell receptors (TCRs) and B cell receptors (BCRs or surface-bound antibodies). Natural killer cells (NK-cells) represent the third type of lymphocyte but are traditionally considered to be part of the innate immune defense due to the lack of antigen-specific surface receptors. However, recently NK-cells have also been shown to present antigen-specific memory [35]. The NK-cells have cytotoxic functions where they kill altered, i.e. tumor or infected, cells [34].

T-cells

The progenitor T-cells express both CD4 and CD8 co-receptors and mature into either naïve CD8+ or CD4+ T-cells. A naïve T-cell (or B-cell) is a mature lymphocyte that has not yet encountered its specific antigen. Naïve CD8+ T-cells become cytotoxic T-cells upon activation by antigens and destroy infected cells through apoptosis (cell-mediated immunity). Naïve CD4+ T-cells will become activated T-helper (Th) cells when i) presented to its specific antigen on a MHC class II on an APC, ii) CD28 protein on the T-cell binds to B7 on the APC, and iii) stimulated by IL-1 and IL-6 released by other activated immune cells. The precursor Th cells differentiate into Th1 if stimulated by IL-12, and to Th2 cells when promoted by IL-4. Moreover, IFN-γ, produced by Th1 cells, stimulates Th1 cell growth and at the same time inhibit Th2. IL-4, produced by Th2 cells, instead inhibit Th1 cells but stimulate Th2 cell expansion [36]. Th1 cells activate macrophages to destroy intracellular microorganisms and activate B-cell production of antibodies of some IgG subclasses, whereas Th2 cells drive B-cell differentiation into plasma cells producing antibodies of all classes (see below) [34].

B-cells

The main roles of B-cells are to produce antibodies against antigens, function as APCs to communicate with T-cells, and to develop into memory B cells. The memory cells help the body to remember a certain pathogen in order to trigger the adaptive immune response quicker upon a second infection by the same pathogen. A naïve B-cell already produces antibodies that are able to recognize a specific antigen, which the B-cell will present on its own MHC class II molecules. If a Th cell recognizes this particular antigen, the CD40 on the B-cell will bind to CD40L expressed on the Th cell. This binding, together with the release of IL-4, IL-5, and IL-6 by the Th cell, will stimulate the activation of the naïve B-cells into either memory B-cells or antibody producing plasma cells (humoral immunity) [34].
Antibodies

Antibodies, or immunoglobulins (Igs), are large Y-shaped proteins consisting of variable regions, which uniquely bind to different molecules on pathogens, and constant regions, which determine the isotype. There are five antibody isotypes: IgM, IgD, IgG (with four subclasses: IgG1, IgG2a, IgG2b and IgG3), IgE, and IgA (with two subclasses: IgA1 and IgA2). The enormous diversity of the variable region of the antibodies (as well as the TCRs) originates from the ability to generate many random combinations of the set of gene segments encoding the variable chains. This is accomplished through the process of somatic rearrangements of variable (V), (diversity (D)), and joining (J) segments followed by the “hyper-mutation” rate in these gene regions [34]. Thus, the antibodies (and TCRs) can recognize and respond to a vast number of pathogens.

All naïve B-cells express surface-bound IgM, which is also the first antibody secreted upon exposure to an infectious agent. IgM is secreted as pentamers and is also the first antibody to be produced by the fetus. Together with IgM, naïve B-cells express smaller amounts of surface bound IgD, which seem to function as a modulator of B-cell selection and thereby shapes the B-cell repertoire [37].

For the B-cell to produce the other antibody isotypes it must undergo isotype-class switching through the recombination of genes encoding the constant regions (also referred to as class-switch recombination (CSR)). The variable region and the antigen specificity remains the same, but the effector function of the antibody changes. Th2 cells induce class-switching to IgG1 and IgE through IL-4-dependent signaling, and to IgG2b and IgA through signaling dependent on the transforming growth factor (TGF)-β. Th2 cells also stimulate IgA secretion of plasma cells after switching through IL-5 secretion. IFNγ, produced by Th1 cells, induce class switching to the remaining two IgG subclasses (IgG3 and IgG2a). IgG is the most abundant of the five antibody types and comprises as much as 15% of the total serum protein. IgG protects the body against pathogens by opsonization, where it binds to the pathogens to help phagocytes to detect them, and agglutination, where IgG bind to many pathogens in order to glue them together into a cluster detectable by phagocytes. IgG also acts as an activator of the complement system, and as a neutralizing antibody by preventing the infectivity of a virus and the toxicity of a toxin by binding to them.

IgE is secreted as a monomer and is the predominant antibody in the tissues where its constant region is bound to mast cells through the high-affinity FcεRI receptors. When IgE cross-links with the antigen, it stimulates the mast cell to release mediators such as histamines. Also, basophils and
eosinophils express the FcεRI and participate in this reaction, resulting in \textit{type I hypersensitivity} (see below) [34].

IgA can be secreted as a monomer or dimer (held together with a J-chain) in the blood, or as a dimer in the mucosal system. IgA is the main immunoglobulin in the mucus and the primary defense against mucosal infections. IgA is most active in the epithelium where it functions as neutralizing antibody. Serum IgA initiate inflammation through interaction with FcαR on immune effector cells [38], which are activated B-cells (plasma cells) or T-cells (cytotoxic or helper T-cells).

**Hypersensitivity reactions**

The immune system is extremely complex and the previous sections have only touched upon parts of the many complicated mechanisms involved in the protection against pathogens. In order to react properly, the immune system has to distinguish between self and non-self, and between pathogenic and harmless substances. \textit{Central immune tolerance} is induced in the thymus and bone marrow, and is the mechanism by which the immune system learns to discriminate self from non-self. An altered central immune tolerance can give rise to autoimmune diseases, when the immune system reacts to certain self-proteins. \textit{Peripheral tolerance occurs} in the lymph nodes to prevent from over-reactivity to environmental entities. If the peripheral tolerance is not established properly, an immune response can be initiated against harmless substances in the environment. These immune responses are referred to as \textit{hypersensitivity reactions}. There are four different classes of hypersensitivity reactions, each characterized by its specific mechanism.

The \textit{Type I hypersensitivity} reaction involves IgE antibody triggering of mast cells. Allergy (including allergic rhinitis, asthma, systemic anaphylaxis, or atopic dermatitis) is often associated with the type I hypersensitivity, also referred to as the immediate-type reactions mediated by IgE in response to \textit{allergens}. Allergens are antigens that do not originate from pathogens and, despite being harmless, initiate immune responses.

The \textit{Type II hypersensitivity} reaction involves IgG antibodies directed against cell surface or matrix antigens and can trigger the complement system. Drug allergy, such as oversensitivity to penicillin, is one example of a type II response.

The \textit{Type III hypersensitivity} reaction are directed towards soluble antigens triggered by antibody complexes. It may result in, for example, serum sickness – the reaction to anti-serum.

The \textit{Type IV hypersensitivity} reaction is T-cell mediated and can be divided into three subgroups. The first group involves the activation of macrophages by Th1 cells to release inflammatory mediators. The second group is
the Th2 cell activation of the inflammatory responses dominated by eosinophils. The third group involves tissue damage caused directly by activated cytotoxic T-cells. Contact dermatitis can be the result of the first and the third group, thus involving either Th1 or cytotoxic T-cell responses. The type IV reaction, involving Th2 response, can lead to chronic stages of asthma, allergic rhinitis and atopic dermatitis [34].

Primary immunodeficiencies

When some parts of the immune system are missing or dysfunctional it results in primary immunodeficiency disorders (PIDs). These do not include immunodeficiencies caused by secondary effects of e.g. another disease, drug treatment, or toxin exposure. In human, there are about 150 different immunodeficiencies identified and more than 120 have been associated with a genetic defect. Humoral primary immunodeficiencies, where one or several antibodies are insufficient, account for 65% of all PIDs. Although selective IgA deficiency is the most frequent PID in Caucasians (1/600 individuals), common variable immunodeficiency (CVID) (1/50,000 in Caucasians) is the most commonly encountered in clinical practice due to the more frequent presentation of symptoms [39]. CVID represent a heterogeneous group of disorders defined primarily by IgG levels below 5 g/l (where no other cause of immune defect is defined) presented at an age over 4 years. CVID patients usually become symptomatic later in life presenting recurrent and/or severe infections, autoimmunity, malignancy, and allergic disorders [40]. About 25% of CVID patients present autoimmune disorders reflecting the failure of preventing auto-reactivity during B-cell development. The majority of CVID patients also show impaired vaccine responses, decreased levels of IgA and/or IgM, and a reduction of memory B-cells [39]. Familial inheritance of either IgA deficiency or CVID has been indicated in about 20% of the cases [41].

Severe combined immunodeficiency (SCID) is a disorder caused by the incapability to generate antigen-specific immune responses and is described in humans, horses, mice, and dogs. The known genetic defects primarily cover single genes, resulting in a failure of lymphocytes to differentiate or defects in the cell surface receptors for ILs on mature lymphocytes [42].

Atopic dermatitis

Atopic dermatitis (AD) is a chronically relapsing and pruritic skin disease caused by inflammation of the epidermis and dermis. In human, AD occurs more often in families with atopic diseases and affects 10-30% of children and 1-3% of adults [43]. In this introductory section of AD, I will primarily
refer to human AD studies (unless mentioned otherwise). However, the immune response in human AD is similar in canine AD (CAD). Next, I will go into specifics about CAD, compare human AD and CAD, and finally discuss environmental and genetic risk factors in both humans and dogs.

The immune response underlying atopic dermatitis

In AD, IgE-mediated activation of mast cells initiates the inflammatory cascade involving eosinophils, basophils, and Th2 cells. Clinical signs of AD include pruritus (itchiness), erythema (redness), and skin lesions such as crusts and lichenification (“elephant skin”), and are caused by the actions of the activated immune cells. Characteristics are also dermatopathological signs of spongiosis (intercellular edema within the epidermis), hyper- and parakeratosis (abnormally thick / nucleated squames) as well as findings of perivascular infiltrations of lymphocytes and eosinophils [44].

The original function of the hypersensitivity reaction type I is to defend the body against some types of parasite infections, such as tick bites on the skin. However, in allergy this inflammatory reaction is an unnecessary response to harmless environmental allergens. In an allergic reaction, mast cell degranulation and Th2 activation cause eosinophils to accumulate and become activated. Activated eosinophils release highly toxic granule proteins and free radicals that kill microorganisms and parasites, but also result in host tissue damage. Basophils have a similar function as the eosinophils and are also recruited to the site of the allergic reaction. The acute inflammatory response is divided into two phases: the immediate and the late-phase reaction. The immediate reaction starts within seconds upon allergen exposure and results in the release of histamine, prostaglandins, proteases, and other pre-formed and rapidly synthesized mediators. These cause increased vascular permeability, dilated blood vessels, contraction of the smooth muscle, and recruitment of inflammatory cells. After about 8-12 hours, the late-phase reaction occurs, caused by the release of additional mediators by the mast cells and the involvement of other cells, including eosinophils and Th2 cells. The acute inflammatory response of AD is characterized by a predominant expression of the Th2-type cytokines IL-4 and IL-13, which assist in the isotype-class switching into IgE, and IL-5, which increases eosinophil survival [45]. In addition, Th2 cells release IL-31, shown to promote pruritus, alopecia (hair loss) and skin lesions in mice [46] and dogs [47].

The level of antigen penetrance through the skin and the antigen capture by immune cells are important features in the context of AD. These are monitored by the skin barrier function, which can be divided into three parts: the stratum corneum (air-liquid barrier), tight junctions (liquid-liquid barrier),
and the Langerhans cell network (immunological barrier). Under the stratum corneum, which is the outermost layer of the epidermis, is the stratum granulosum [48]. The tight junctions are positioned in the middle section of the stratum granulosum and functions as a “gateway” for water, ions and solutes passing through the intercellular space between the cells in the epithelium [49]. Langerhans cells (LCs) are specialized APC dendritic cells residing in the skin that can be activated by perturbation to the stratum corneum. The activated LC can extend its dendrites through the tight junction barrier to efficiently take up allergens, which they will present through their MHC molecules to the T-cells [50]. LCs are responsible for predominantly priming the Th2 cells [51]. This LC-mediated Th2 response is further intensified by the thymic stromal lymphopoietin (TSLP), released by keratinocytes upon detection of pathogen skin intrusion [52]. Moreover, IgE receptors are also expressed on LCs and the IgE receptor cross-linking on LCs as well as microbial products (e.g. staphylococcal enterotoxins) contribute to the recruitment of inflammatory dendritic epidermal cells (IDECs) into the skin. IDECs are mainly detected in inflamed skin and have a high expression of FcεRIs on their surfaces, thus providing an efficient uptake of allergens through IgE. Allergen binding of IDECs leads to the production of IL-12 and IL-18, which may contribute to a switch from the initial Th2 response to the predominance of IFNγ-producing Th1 cells. This seems to be a crucial step for the cornification of the skin lesions [51, 53]. IFN-γ and IL-12 have been detected in chronic phase AD skin from both dogs [54] and human [53]. Therefore, AD can be described as a biphasic inflammation, where there is a switch from the Th2 response in the acute inflammatory response of AD to a Th1 response in the chronic stage of AD skin inflammation.

Canine atopic dermatitis

Similarly to human AD, CAD is defined as an inflammatory and pruritic allergic skin disease caused by both genetic and environmental factors [55, 56]. The overall prevalence of CAD is estimated to 3-15% in studied dog populations [57, 58] and is highly overrepresented in certain dog breeds, which strongly indicates a genetic background. Previous clinical and epidemiological studies have defined predisposed breeds including Boxer, Bullterrier, West Highland white terrier, German shepherd, and Labrador retriever [59-62].

Clinical signs

The clinical signs of CAD, appearing as pruritus and an eczematous skin, is most commonly associated with IgE antibodies directed towards environmental allergens [63]. Face, ears (otitis externa), paws, extremities, ventrum, and flex-zones are typically affected by pruritus and erythema [64]. The
affected body regions seem to differ between breeds, for example the German shepherds were found typically affected by otitis externa and eczema in the ventrum, whereas the French bulldogs and Boxers showed a predisposition for symptoms in the face [61].

Diagnosis
The diagnosis of CAD is established after an extensive investigation [63], where other possible causes of the clinical presentations are carefully excluded. These other conditions include scabies or other pruritic ectoparasite infestations, pruritic bacterial skin infections, *Malassezia* (yeast) dermatitis, flea allergy dermatitis, but also the more rarely occurring cornification disorders and contact dermatitis. In addition, the presence of *cutaneous adverse food reactions* (CAFR) needs to be evaluated through a strict diet trial in order to establish the CAD diagnosis. The reason is that CAFR can be caused by either a hypersensitivity reaction or have a non-immunological cause, but appear clinically identical to CAD. Thus, if the symptoms are caused only by CAFR, a CAD diagnosis could be incorrect. Finally, a positive allergen-specific IgE test (serology or intradermal test) concludes the CAD diagnosis and may define the offending allergens.

The reliability of the IgE testing and the correlation between intradermal- and serology allergen-specific IgE tests have been widely discussed [65-67]. For instance, the proportion of CAD-affected dogs with a positive IgE test result (i.e. sensitivity) was reported to reach 53.6%, and the proportion of healthy dogs with a negative test result (i.e. specificity) was 84.4% [66]. Thus, a positive allergen test alone is not conclusive for establishing a CAD diagnosis. Moreover, it has been shown that the food-specific serum IgE reactivity measurements do not correlate between laboratories or with the disease status, thereby making the food-allergen IgE testing unreliable in evaluating CAFR [68].

Treatment
The severity and clinical presentations of CAD differ between patients and the most optimal treatments may also vary from patient to patient. The primary action is to avoid the offending allergen and use frequent bathing with emollient shampoos to try keeping the clinical symptoms under control. If allergen avoidance is not feasible and the clinical signs are not reduced sufficiently, the use of anti-inflammatory pharmacotherapy is necessary. These anti-allergic drugs include cyclosporin A (primarily preventing mast cell degranulation), anti-histamines (prevent the effects of histamine), and glucocorticoids (inhibitor of cells activated in the response to allergens). Also, the use of antimicrobial interventions may be necessary to prevent and treat secondary infections, e.g. by *Malassezia* (Figure 3) or *Staphylococcus*, in
atopic dogs. Allergen-specific immunotherapy (ASIT) may be successful and relevant in dogs where allergen avoidance is impossible and anti-allergic drugs are ineffective [69].

Figure 3. Secondary infections caused by *Malassezia* resulting in hyperpigmentation and thickening of the skin in two CAD affected dogs: a West Highland white terrier and a German shepherd dog, respectively. The pictures were kindly provided by K. Bergvall.

Comparisons between canine and human atopic dermatitis

The disease manifestations of CAD and human AD show high resemblances [70, 71]. The similarities include clinical signs (Figure 4), histological signatures, immune response patterns, age of onset, genetic predispositions, and prevalence (summarized in Table 1). The main dissimilarity between the two species is that humans often experience the atopic march, which is when the eczema in a child progress into asthma and allergic rhinitis in the adult patient [72]. More than 60% of the children with AD are at risk to exhibit the atopic march [73]. This phenomenon is not seen in dogs. Instead the eczematous symptoms remain throughout the dogs’ lives [74], and they are typically not affected by asthma. Despite the likenesses of the disease presentations, the skin structure differs morphologically between human and dog. The most obvious difference is the hair growth, which in the dog grows in bundles and in cycles as opposed to the solitary and more continuous growth in human. The sweat glands in the dog are fewer than in human and have other primary functions, such as in secreting pheromones and help sealing the outer layer of epidermis. In addition, the epidermis in the dog is much thinner (3-5 cells thick) compared to human (10-15 cells) [75].
Figure 4. Atopic symptoms including pruritus and erythema presented similarly in the flexure areas in a Chinese crested dog and a human baby. The pictures were kindly provided by K. Bergvall.
Table 1. Comparison of canine and human AD. Modified from Marsella and Girolomini, review 2009 [76].

<table>
<thead>
<tr>
<th></th>
<th>Canine AD</th>
<th>Human AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence in the general population (%)</td>
<td>10-15</td>
<td>5-20 of children</td>
</tr>
<tr>
<td>Genetically inherited</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Age of onset (years)</td>
<td>1-3</td>
<td>&lt;1-5</td>
</tr>
<tr>
<td>Skin areas affected</td>
<td>Face, skin folds</td>
<td>Face, skin folds</td>
</tr>
<tr>
<td>Spongiotic dermatitis (characteristic eczema)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Skin-infiltrating eosinophils</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Skin infiltration by IgE+CD1c+ dendritic cells (e.g. LCs)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pruritus</td>
<td>Severe</td>
<td>Severe</td>
</tr>
<tr>
<td>Skin xerosis (dry skin)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Increased TEWL (water loss)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Decreased epidermal filaggrin</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Higher skin colonization by <em>Staphylococcus aureus</em></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Th2-dominated immune responses</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Atopy patch test</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>IgE-specific responses (%)</td>
<td>80</td>
<td>55-90</td>
</tr>
<tr>
<td>Rhinitis and conjunctivitis (%)</td>
<td>&lt;5</td>
<td>35</td>
</tr>
<tr>
<td>Asthma (%)</td>
<td>&lt;5</td>
<td>30</td>
</tr>
<tr>
<td>Atopic march</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

In humans, there is a clear indication of allergen sensitization based on the total serum IgE levels. In AD patients the total IgE levels can be up to 1.5 μg/ml, compared to non-atopic humans having IgE levels of approximately 0.15 μg/ml. In contrast, IgE levels in dogs do not relate to CAD status and the total levels of IgE (ranging from 1 to 41 μg/ml) are about 100 times higher than in non-atopic humans. Thus, total IgE levels cannot be used as a diagnostic tool for CAD. Interestingly, the IgE levels in juvenile dogs are more similar to human levels (0.2-1.0 μg/ml). The extraordinarily high IgE levels in adult dogs may be explained by the higher exposure to a parasite-rich environment, in contrast to what is common for humans and young dogs [77].

Notably, other species that have documented cases of AD, including horses [78] and cats [79], are also featuring a breed variability suggesting a genetic background. Overall, allergic responses are seen in all of the domestic animals, rodents, non-human primates, as well as in avian species [80]. Howev-
er, the dog is the only known animal that naturally develops skin lesions similar to human AD [81].

Environmental factors

The increased prevalence of both CAD and human AD over the last decades has generated many discussions involving the influence of environmental factors. It has been proposed that a new life style, basically defined by a cleaner environment for both human and dogs, increases the risk for developing this disease. Likely, the lack of early exposure to various infectious agents, parasites, as well as harmless microorganisms, leads to an impaired establishment of immune tolerance.

The prevalence of AD in the overall human population is 2-5% [82] and has presented with an increasing trend, which now seems to have reached a plateau around 20% in countries with the highest prevalence e.g. Nigeria, the United Kingdom, and New Zealand [83]. Environmental demographic risk factors include higher socioeconomic status, family education, urban environment, and small family size [83]. A protective effect has been seen with a generally high exposure to non-pathogenic microbes through, for example, early daycare, unpasteurized farm milk, and animal exposure [84]. This phenomenon is referred to as the hygiene hypothesis and was proposed already in 1989 when the number of children in a household was negatively correlated to AD development [85]. Moreover, a high maternal intake of dairy products was shown to reduce the risk for AD in humans [86]. The hygiene hypothesis has had further support [87], but opposing evidence has also been reported [88]. Somewhat contradictory is the high prevalence of AD in Nigeria, where heat was implicated as the most aggravating factor [89].

Most of the factors that are linked to an increased incidence of AD in human are also consistent with the change of the dog’s environment the last decades, including more time spent indoors in a pathogen-reduced environment [57]. Indeed, studies in dogs support a hygiene hypothesis also in CAD development. For example, maternal feeding of non-sterilized diet during lactation increased the risk for CAD in the puppies [90], a higher prevalence of CAD in dogs living in cities in Sweden [59], and a decreased risk for dogs living in the countryside and/or together with other dogs or cats [55].

Genetics

Besides being influenced by environmental factors, AD in humans and dogs are primarily affecting individuals with a genetic predisposition. Several
genetic risk factors have been suggested to contribute to the development of AD in human and dogs (reviewed in [91] and [92]).

**Human**

AD is a complex disease caused by multiple genetic and environmental risk factors. The strong familiar predisposition [93] suggests that genetic factors play a substantial role in the disease development. Twin studies confirm the role of genetics in the development of AD based on twin concordance rates (*i.e.* the probability that both individuals in a pair of twins are affected by, in this case, AD) up to 86% for monozygotic twins and 50% for dizygotic twins [94]. In a recent large study, including ~30,000 individuals from Europe and Asia, 11 susceptibility loci explained 14.4% of the total heritability for AD [95].

The genes associated with AD seem to represent two main pathophysiological groups: *i)* immune-mediated pathways and *ii)* skin barrier functions [91]. To date, six GWAS (or comparable studies) have been reported for AD presenting several candidate genes. Many of the findings have been replicated in additional studies [91] including the gene *C11orf30* (chromosome 11 open reading frame 30) encoding a nuclear protein shown to regulate cytokine responses [96]. A large meta-analysis, including 16 population-based cohorts, presented significant association with SNPs near the genes *OVOL1* (ovo-like zinc finger 1) and *ACTL9* (actin-like 9) involved in epidermal proliferation and differentiation, and within the gene *KIF3A* (kinesin family member 3A) located inside a cytokine gene cluster [97]. There have also been reported associations to the MHC gene region [91]. However, most striking are the associations of AD to the *FLG* gene reported in more than 20 studies [98]. The *FLG* gene encodes the protein filaggrin, which is important for maintaining a protective skin barrier as it takes part in facilitating terminal differentiation of the epidermis. Filaggrin aggregates keratin filaments into tight bundles, promoting the collapse of epidermal cells into flattened squames in the stratum corneum. Filaggrin and keratin constitutes 80-90% of the cornified envelope, which is an insoluble protein structure consisting of flattened squames without nucleus. The cornified envelope is surrounded by lipids [99] and important for the protection against environmental agents and epidermal water loss. *FLG* loss-of-function mutations were first identified to cause ichthyosis vulgaris, another chronic skin disease. Subsequently, the mutations were found to strongly associate with AD [100]. Reported frequencies of *FLG* mutations are between 18 to 48% in AD patients [101] compared to an average frequency of around 9% in Europeans [100]. To date there are approximately 40 *FLG* loss-of-function mutations reported with frequencies differing greatly between human populations [102].
There are additional AD-associated genes affecting the skin barrier. One is the *CLDN1* gene encoding claudin-1 important for the permeability function of the tight junctions [49]. Moreover, the gene *SPINK5* has shown association to AD across ethnicities [103]. The encoded protein, serine protease inhibitor of Kazal-type 5, plays a role in maintaining homeostatic desquamation (i.e. the shedding of dead corneocytes from the stratum corneum) and a defective protein gives rise to increased desquamation, dry skin, hair abnormalities, and AD [104]. Despite many associations to important genes, much of the genetics underlying AD in humans remains to be explained.

**Dog**

The striking breed predisposition of CAD [59-62] and heritability estimates, as high as 47% (+/-17) in populations of Labrador and Golden retrievers [105], strongly supports a genetic basis also for CAD. Previous genetic studies in dogs have revealed associations between CAD and: i) a locus on chromosome 17 in West Highland white terrier breed [106, 107] harboring the top candidate gene *PTPN22* (protein tyrosine phosphatase, non-receptor type 22) involved in the inhibition of spontaneous activation of T and B-cells, ii) two intergenic SNPs in an eight breed meta-analysis [108], and iii) one SNP in the gene *TSLR* (thymic stromal lymphopoeitin receptor) in an eight-breed candidate gene study [109]. Interestingly, an altered mRNA expression of filaggrin was detected in atopic dog skin when compared to healthy control skin, indicating the involvement of the *FLG* gene also in CAD development [110]. Despite these results, which are suggesting both immune-related pathways and skin barrier involved in CAD likewise to human AD, much of the genetic risk factors for CAD still remain unknown. In order to gain statistical power, previous studies have included dogs from multiple breeds. However, considering the heterogeneity and breed differences this may not be sufficient as the only approach. Most likely different breeds harbor, at least partly, different genetic alterations.

**Immunoglobulin A deficiency**

IgA is the key antibody in the mucosal defense and the second most abundant antibody in human serum, where it initiates effector functions of the immune system [111, 112]. Concentrations in human serum are around 2-3 g/l compared to the most prevalent antibody IgG, which normally present at levels of around 12 g/l. Serum IgA metabolizes approximately five times faster than IgG, thus the production rates should be about the same. Serum IgA is primarily monomeric whereas secretory IgA comprises mainly of dimeric forms and serves a variety of functions to protect the mucosal surfaces of the respiratory, gastrointestinal, and genitourinary tracts. These are
substantial tissue areas with potential exposure of inhaled or ingested pathogens [112]. IgA has been detected in serum, feces, and tears, but also in the sweat glands [75, 113], suggesting that IgA functions also as a cutaneous secretory immunoglobulin.

**Human IgA deficiency**

Most patients with selective IgA deficiency are asymptomatic but approximately one out of three suffers from recurrent infections at mucosal areas and/or autoimmune diseases [111]. Also allergy, including AD, is presented more commonly in IgA deficient patients [114]. A recent study proposed a more potent effect of IgA deficiency than what has been generally accepted, owing to a substantial increase in allergic diseases, autoimmunity, pneumonia, and susceptibility to infections in IgA deficient patients [115]. Moreover, a significantly reduced secretion of IgA was detected in skin [113] and in tears [116] from AD patients compared to healthy controls.

**Diagnosis**

Selective IgA deficiency is defined by a serum IgA concentration below 0.07 g/l, combined with normal levels of IgG and IgM in individuals older than four years of age [117].

**Genetics**

The MHC region has been associated with IgA deficiency and both protective and risk alleles have been reported. The most common haplotype found in IgA deficient patients was also linked to various autoimmune diseases with known association to IgA deficiency [118]. Moreover, non-MHC genes have been associated with IgA deficiency. Well-studied examples are the IFIH1 gene (interferon-induced helicase 1) and CLEC16A gene (c-type lectin domain family 16, member A). Both genes have also been associated to autoimmune diseases, again suggesting that autoimmune mechanisms may contribute to the pathogenesis of IgA deficiency [119] or possibly that causative genes have a pleiotropic effect on IgA deficiency and autoimmunity in separate pathways.

The α1 and α2 genes, encoding the heavy and the light chains of IgA, are normally expressed and functional in IgA deficient patients and no coding mutations have been detected in these genes [120]. Thus, the deficiency is likely rather caused by a defective regulation of the IgA class switching and/or secretion, or possibly by aberrant levels of ILs.
Low IgA levels in dogs

There are genetically and experimentally induced rodent models produced to study IgA deficiency. However, none of them really resembles the human disease from a clinical perspective. Spontaneous models for IgA deficiency are reported in both dog [41] and chicken (affected by autoimmune thyroiditis) [121], but here the underlying molecular basis is not fully understood. Low IgA levels have been frequently reported in German shepherds [122-126], but also in Shar-Pei [127] and selected populations of Beagles [128, 129], compared to other dog breeds. Dogs with low serum IgA levels clinically resemble human IgA deficiency when it comes to being affected by recurrent infections [127, 128, 130] and immune-mediated diseases [131]. Despite many studies, there is no general cut-off to distinguish normal from abnormal IgA, IgG, or IgM levels in dogs, whereas the reports rather propose breed-specific values [128, 132, 133]. Hence, no diagnostic criteria have been established to define canine IgA deficiency. Therefore, the use of the dog as a model for IgA deficiency should be based on defined low IgA level breeds, which differs from the general dog population, and the fact that the clinical phenotype resembles human IgA deficiency.
Aims of the thesis

The overall aim of this thesis was to discover genetic risk factors for atopic dermatitis and IgA deficiency by using the dog as a genetic model.

The specific aims were:
- To identify genetic loci associated with CAD in dogs.
- To identify dog breeds predisposed to low IgA levels.
- To identify genetic risk loci for low IgA levels in dogs.
- To identify candidate genes and biological pathways for future research on atopic dermatitis and IgA deficiency in dogs and humans.
Present Investigations

Papers I & II: CAD-associated genetic variants are located in tissue-specific enhancers within the PKP2-locus and participate in transcriptional up-regulation

Background

Atopic dermatitis
Humans and dogs are both affected by the allergic skin disease atopic dermatitis (AD), caused by interactions between genetic and environmental risk factors. The immune responses in AD patients are primarily due to immunoglobulin E (IgE) antibodies recognizing harmless environmental allergens, resulting in a degranulation of active mediators, such as histamine, by mast cells and eosinophils. There is an early onset in both species and typical signs include pruritus and inflammation of the skin, most commonly localized to flex and friction zones. AD can be described as a biphasic inflammation as the expression of Th2-type interleukins (IL-4, IL-5, and IL-13) dominates in the acute AD phase, whereas Th1 cells, producing cytokines including IFN-γ and IL-12, are elevated in the skin of patients affected by chronic AD [53]. Furthermore, an impaired skin barrier function may enhance allergen penetration and subsequently start the cutaneous inflammation driven by the Th2 cells. The skin barrier may be altered due to genetic alterations (e.g. mutations in the FLG gene) or physical stress (e.g. heat, sweat, or dryness).

German shepherds
In dogs, there are reported breed predispositions to canine AD (CAD), and the German shepherd dog (GSD) is one of the commonly affected breeds [59-62]. GSDs show a high susceptibility also to other immunological disorders, including circumanal fistulae and pancreatic acinar atrophy [62, 134, 135]. In a Swedish epidemiological study from 2013, the GSD was the most over-represented breed for immunological disease (a relative risk (RR) of 2.7) and displayed a mortality RR due to the general category “skin problems” of 7.8, when compared to all other breeds combined [62]. In addition, low serum IgA levels have frequently been reported in GSDs [122-124, 126]. Considering the number of registered dogs, the GSD is the most com-
mon breed in Sweden (www.jordbruksverket.se, statistics 2014). This enables collection from a large number of samples, at least 100 cases and 100 controls, required for GWAS of a complex trait based on power calculations, assuming a 5-fold risk [17]. However, other dog GWAS of complex diseases have identified associated loci by using fewer dogs, such as for canine SLE-like disease [33], osteosarcoma [20], and cardiomyopathy [136]. Ideally the sample cohort in a GWAS should be as unrelated as possible. However, this is practically difficult to achieve fully within a dog breed. Yet, pairs of control-control or case-case siblings should be avoided to limit the risk for confounding effects.

**Regulatory mutations**

For complex traits such as CAD, we do not expect to find protein-coding mutations in single genes with complete association to the trait. Rather, we presume to find mutations that are affecting gene expression in certain tissues and conferring a higher risk of the disease. There is evidence that the strong selection in domestic animals has resulted in the accumulation of multiple causal mutations affecting the regulation of gene transcription [137]. This is exemplified in the genetic mapping of the white coat color phenotype in dogs, where different combinations of causal polymorphisms resulted in haplotype effects, causing an altered transcription of the MITF gene [138].

**Paper I: Genome-wide association of a locus on chromosome 27 with CAD in German shepherds**

**Study design**

We collected blood and serum samples from CAD cases and from healthy controls representing a total of 207 GSDs. The aim was to identify genomic region(s) associated to CAD in GSDs by using GWAS, and to evaluate whether or not IgA levels were correlated to CAD in GSDs.

**CAD inclusion criteria**

The CAD diagnoses, defining the cases, were established by first ruling out other possible causes to the clinical signs, such as ectoparasite infestations, *Staphylococcus* pyoderma and *Malassezia* dermatitis. In addition, a hypoallergenic dietary trial of at least 6-8 weeks, followed by a challenge period, was conducted to evaluate if CAFR was entirely or partially causing the symptoms. CAD was concluded if the dog was not adequately controlled on the hypoallergenic diet and had positive reactions on intradermal allergy test or IgE serology test. The healthy controls were over five years of age without immune-related disease or any problems with pruritus, repeated ear infections, or skin lesions compatible to the clinical signs of CAD. The classi-
fications of controls were based on owner questionnaires and/or clinical examinations. In addition, control dogs were excluded as CAD controls in the final analysis if the serum IgA levels measured 0.10g/l or lower.

IgA measurements
The serum samples were measured for IgA concentrations with enzyme-linked immunosorbent assay (ELISA). This method of quantifying IgA concentration can briefly be described in five steps: i) the wells of the assay plate were coated by capture antibodies directed against dog IgA (goat anti-dog IgA antibodies), ii) unoccupied areas of the wells were blocked to prevent unspecific binding by using milk powder as a blocking agent, iii) diluted serum samples were added and the capture antibody bound to the IgA in the serum, iv) a secondary enzyme-linked antibody (AP-conjugated goat anti-dog IgA) was added that bound to the antigen (i.e. the dog IgA), v) finally, a substrate was added that was converted by the enzyme (linked to the secondary antibody) to become detectable and quantified.

Analyses and results
First, we examined the relationships between phenotypes and other variables. We detected that CAD cases had significantly lower IgA levels compared to the controls (p=1.1x10^{-5}) and that the age of the dog at the time of sampling, i.e. ‘age at sampling’, was positively correlated to IgA levels.

Population substructure
We genotyped in total 207 GSDs with the Illumina 170K Canine BeadChip and the initial analysis revealed a high genomic inflation factor (λ) of 1.3, indicating population stratification. By studying the MDS-plot of genetic relationship in the study population, we noted a formation of two subpopulations with an uneven distribution of CAD cases and controls. Indeed, there was a pronounced difference between the subpopulations in risk of developing CAD (OR=4.4, CI95=2.3-8.8). When comparing the merits of the GSDs we noted that the division into two subpopulations was most likely due to the selection for a working type GSD and a show type GSD. We concluded that the GSDs bred for working capabilities carried a lower risk for developing CAD compared to the GSDs bred for the phenotype preferred in dog shows. To correct for the subpopulations as well as cryptic relatedness we used a mixed model approach in the final association analysis. In addition, IgA levels and ‘age at sampling’ were included as covariates. Dogs excluded from the analyses were: i) lacking CAD and/or IgA status, ii) CAD controls with low IgA levels, iii) apparent genetic outliers, and/or iv) too related.
Association to a locus chromosome 27 harboring the PKP2 gene

We analyzed in total 91 CAD cases and 88 healthy controls in the final GWAS and detected a genome-wide significant association to a locus on chromosome 27. The associated locus spanned ~1.5Mb and consisted of 21 SNPs in LD ($r^2>0.8$). The region contained eight genes and the top two associated SNPs surrounded the *PKP2* gene, encoding the protein Plakophilin-2 (PKP2). By performing targeted re-sequencing of the associated region, we attempted to identify all variants associated with the phenotype. We aimed at defining potential causative variants and evaluating if any of these were more associated than the GWAS top SNP. We sequenced one case that was homozygous and two cases that were heterozygous for the risk haplotype, and two controls lacking the risk haplotype (i.e. homozygous for the control haplotype). Within the entire associated regions, ~2,500 SNPs followed the pattern of the risk and control alleles of the top GWAS SNPs. We selected in total 54 SNPs for further genotyping. Most of the selected SNPs were located within the *PKP2* gene region, primarily in conserved elements but also in non-conserved regions. A few of the SNPs were selected to sufficiently cover the whole candidate region and we also included the two top GWAS SNPs. The SNPs were genotyped in the same sample set used for GWAS and 42 SNPs remained for analysis after quality controls. In an attempt to define more distinctly separated haplotype formations with strong associations in the GSD, we assigned SNPs to blocks based on their LD pattern ($r^2\geq0.9$). The two most associated blocks were positioned directly upstream and downstream of *PKP2*. The association analysis of the single fine-map SNPs defined three top SNPs, which were equally associated as the top blocks. One of the three SNPs was the top GWAS SNP, thus no SNP with greater association compared to the GWAS could be defined. Based on these results we defined a ~209kb associated region, including the *PKP2* gene, that likely harbors the causative variants predisposing GSDs to CAD.

Paper II: Four CAD-associated SNPs within the *PKP2*-locus show regulatory effects on gene transcription

**Study design**

After our initial fine-mapping of the locus on chromosome 27, we opted to perform a more dense genotyping in the associated region to evaluate if a shorter and more strongly associated haplotype could be identified. In addition, we aimed at pinpointing the causative variants and evaluate their functionality.
Samples and SNPs
Since multiple breeds are predisposed to CAD, we aimed to evaluate whether or not other breeds share the same genetic risk factor as the GSD. Therefore we included cases and controls from five additional high-risk CAD (HRCAD) breeds. While expecting that low-risk CAD (LRCAD) breeds do not share the risk variants, we included samples from five LRCAD breeds as controls. The entire sample set consisted of 370 dogs including 174 GSDs (the same set of dogs as were used in GWAS except for five controls), five HRCAD breeds, five LRCAD breeds, plus 1 or 2 samples from four more breeds (Table 2). We genotyped 120 SNPs, selected i) to match the associated haplotype, ii) to cover the associated blocks and the nearby region, and iii) based on its presence in other breeds and species. After quality controls, 104 SNPs remained for the association analyses in GSDs only, in all breeds together, and in each HRCAD breed separately.

Table 2. Dogs included in the genotyping of 120 SNPs.

<table>
<thead>
<tr>
<th>Breed</th>
<th>CAD cases</th>
<th>CAD controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSD*</td>
<td>91</td>
<td>83</td>
</tr>
<tr>
<td>Labrador retriever*</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td>Golden retriever*</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>West Highland white terrier*</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>Boxer*</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>Bull terrier*</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>Irish soft coated weaten terrier*</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Jack Russell terrier*</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Wachtelhund*</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Elkhound#</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>Howavart#</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Giant Schnauzer#</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>Smalands hound#</td>
<td>-</td>
<td>9</td>
</tr>
<tr>
<td>Irish wolfhound#</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Mixed</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total number of dogs</strong></td>
<td><strong>178</strong></td>
<td><strong>192</strong></td>
</tr>
</tbody>
</table>

* HRCAD breeds, # LRCAD breeds.
** Number after quality controls.

Luciferase reporter assays
The candidate variants that were selected based on the association analyses were functionally evaluated using luciferase reporter assays. The DNA fragments harboring the SNPs were cloned in front of the minimal promoter in the pGL4.26 luciferase reporter vector. The plasmids were validated by
sequencing and purified before transfection into the four following cell lines: *Madin Darby Canine Kidney epithelial cells* (MDCK), *human immortalized keratinocytes* (HaCaT), *human T cell leukemia line* (Jurkat), and *human erythromyeloblastoid leukemia cell line* (K562).

**Analyses and Results**

**GSD association analysis**

Two top associated SNPs (one being the same as the top GWAS SNP) were defined in the GSD association analysis, and the associated haplotype, consisted of 11 highly linked and associated SNPs, spanned ~282kb. Thus, we were unable to narrow down the region compared with the previous fine-map, but rather expanded it through the use of a denser SNP set.

**All breeds association suggests the candidate SNP 27:19093355**

In the *all breeds* association analysis, we identified an associated haplotype consisting of seven SNPs spanning intron 6 of *PKP2* across the 3'UTR and 19kb downstream of *PKP2*. The risk haplotype was present in 37.3% of the GSD cases and 9.6% of the GSD controls and was carried also by Labrador retrievers (LRs) and by one Golden retriever (GR), but not detected in any of the other breeds. Half of the haplotype was shared by West Highland white terriers (WHWTs). Despite being rare, the risk haplotype conferred an OR in GSDs of 5.7 compared to 4.1 for the top GWAS SNP. Using the re-sequence data from Paper I, we searched for candidate variants based on the risk/control haplotype pattern and the presence of the risk alleles in HRCAD breeds, human, or other species, and identified nine candidates.

Next, we screened these variants in a transcription factor prediction database (online web tool TRAP [139]), and the most significant prediction was detected for SNP 27:19093355, based on the binding of members of the GATA family of transcription factors to the risk but not to the wild-type allele. The transcription factor GATA-3 is known to stimulate IL-4 production and thereby precursor Th cells to differentiate into Th2 cells. Through additional genotyping of the SNP 27:19093355 in 43 other breeds, we defined eight other breeds, typically LRCAD breeds, as carriers of the risk allele. Genotyping of the SNP 27:19093355 in additional samples from HRCAD breeds showed that the risk allele was present with a high frequency in LRs and WHWTs, but without association to CAD, and at a very low frequency (1%) in GRs. In an additional cohort of GSDs the SNP 27:19093355 conferred an OR of 2.7 for CAD. This was lower than in the original GWAS GSD cohort (OR=6.0).
Four SNPs for functional evaluation
The final list of candidate variants included the two top SNPs from the GSD association analysis: SNP 27:19086778 and SNP 27:19140837 (also top GWAS SNP [140]), the SNP 27:19093355 from the all breeds association analysis, and the second top GWAS SNP 27:18861228 [140]. We phased the four SNPs and defined two risk haplotypes and one common control haplotype in GSDs. One risk haplotype consisted of the risk alleles at all four loci, whereas the other did not carry the rare risk allele at SNP 27:19093355. In GSDs, 61.5% of the cases carried risk haplotype(s) compared to 29.8% of the controls (p=3.7x10^{-5}, OR=3.8). The risk haplotypes were also detected in WHWTs, LRs, and GRs.

The regulatory potential of the top associated SNPs was evaluated using luciferase reporter assays, and the transfections were performed in four different cell lines (mentioned above) representing epithelial and immune cells that might be relevant to the development of the disease. We found that all four regions had regulatory potential on the transcription of the luciferase gene in at least one of the tested cell lines. In the cell lines where allelic differences were observed, which differed between the regions, the risk alleles were consistently associated with an up-regulation of the reporter. There was a very strong induction by the fragment that included SNP 27:19086778 in the skin-relevant cell lines HaCaT and MDCK. This is in agreement with the strong ENCODE enhancer-associated marks active in keratinocytes and epithelial cell lines within the same genomic region in the human genome.

Discussion

Genome-wide association
In GSDs, we associated a ~1.5Mb genomic region to CAD by using GWAS. Through fine-mapping, we identified two top associated SNPs and defined a ~282kb associated region harboring PKP2 as the single gene. However, due to the extensive LD we were unable to narrow down the associated haplotype any further using only GSDs. Instead, we performed an across-breed analysis that outlined a shorter risk haplotype positioned in the GSD associated haplotype within the PKP2 gene, overlapping with signals of regulatory potential in skin-related cell types.

Breed-specific risk factors
The complex nature of CAD where many genetic risk factors, some of which are breed-specific, likely act together may explain why the risk haplotype was not associated in LRs or WHWTs. Possibly, this risk haplotype is merely a modifying locus, additional to a major locus in these breeds. Thus, the PKP2-risk haplotype may solely predispose the whole breed to CAD and the
final trigger for an individual to become atopic may be due to another hitherto unknown major risk factor. For the WHWTs, the low number of cases and controls, and thereby the lack of statistical power, may be another reason for the lack of association.

**Candidate genes**

The four different regions had different effect on gene transcription in the tested cell lines. These findings are in line with the *multiple enhancer variant* hypothesis, stating that several SNPs in LD may influence multiple enhancers for a gene [141], each with a modest effect on gene expression in specific cell types. The observed enhancers, tagged by the top four SNPs, related both to epithelial and immune cells and possibly exert their effects independently in certain cell types. A recent report shows that enhancer regions in dogs may be separated from the affected gene by several non-relevant genes and by over 1Mb [142]. Thus, despite the location of the SNPs, close to and within the *PKP2* gene, we are, at this point, unable to conclude whether transcription of *PKP2* or any of the neighboring genes: *DNM1L*, *YARS2*, or *FGD4*, are altered due to the risk variants. However, the position of the SNPs, especially SNP 27:19086778, which is located close to a keratinocyte-specific enhancer region inside the *PKP2* gene, suggests that *PKP2* transcription might be regulated by this enhancer in skin cells in the first place.

*PKP2* proteins are primarily recognized for their functions in the *desmosomes*. Desmosomes are intercellular mechanical junctions that provide force transmission between cells through their linkage of *intermediate filaments* (IFs) to cell membranes, and contribute to strength and integrity in tissues that exhibit mechanical stress, such as the myocardium and epidermis [143]. Desmosomal proteins come from the three major gene families of *cadherins*, *armadillo proteins*, and *plakins* (Figure 5). Plakophilins belong to the armadillo protein family and are crucial in recruiting the *desmoplakin* (DP) to the desmosomal plaque. The DP provides the important linkage to the stabilizing and stress-bearing IFs inside the cell [144]. The patterns of expression differ between *PKP1*, 2, and 3, where *PKP2* expression is restricted to the bottom layers of the epidermis (*stratum basale* and *stratum spinosum*), whereas *PKP3* is expressed more evenly in all layers and *PKP1* is mostly expressed in the top layers of the skin [145]. The *stratum basale* consists primarily of basal keratinocyte cells (*i.e.* the “stem cells” of the epidermis), which later migrate upwards in the epidermis to finally lose their nucleus and become flattened keratinocytes in the *stratum corneum* [146].
Figure 5. A schematic representation of a desmosome with its major desmosomal proteins. Desmoglein (Dsg) and desmocollin (Dsc) are transmembrane members of the cadherin family that form the adhesive border between the cells. The cytoplasmic tails of the cadherins bind to the plakoglobin (Pg) and plakophilin (PKP), members of the armadillo family. Finally, the plakin family member, desmplakin (DP) links the intermediate filament (IF) to the desmosome. (The picture was modified from Green et al., 2007 [144]).

Despite the absence of mRNA or protein expression data from appropriate tissues in GSDs, we can use the functional data presented in this study to propose functionality for the SNPs and their potential role in the development of CAD in the GSD breed. However, additional biological samples from affected and healthy control dogs (e.g. skin biopsies and/or peripheral blood mononuclear cells, PBMCs) are necessary to further evaluate the functional effect of the variants in GSDs. Moreover, we need to determine whether or not additional cell types are affected by this locus, and if other regulatory variants, located in these or additional enhancer regions, are involved in the gene regulation.
Papers III & IV: Genome-wide analyses suggest novel genes involved in the regulation of serum IgA levels in German shepherds and Shar-Pei

Background

IgA levels in dogs

Previous studies have reported low IgA levels in GSDs [122-126], Shar-Pei dogs [127], and selected populations of Beagle dogs [128, 129]. Dogs with low IgA levels are susceptible to recurrent infections [127, 128, 130] and immune-mediated diseases [131] in concordance to the clinical signs presented in humans with IgA deficiency. Despite many attempts no diagnostic criteria have been established to define IgA deficiency in dogs, reports rather propose breed-specific values. Suggested cut-offs to define canine IgA deficiency mark the lower limit for the 95% CI of the mean in the studied populations [128, 132, 133]. However, these thresholds are not physiologically proven to distinguish between normal IgA and IgA deficiency.

The low IgA levels reported in the feces of GSDs [126] may correlate with the increased susceptibility to enteropathies (i.e. pathology/disease in the intestine) seen in the breed. Indeed, IgA production in the intestine was significantly lower in GSDs with small intestinal enteropathies compared with dogs of other breeds with the same type of disorders, suggesting that a deficiency or dysregulation of IgA at the local mucosal sites may be contributing to small intestinal diseases in GSDs. The same study showed no evidence for IgA deficiency in unaffected mucosal sites or in serum [147]. Interestingly, secreted IgA in tears and saliva, but not serum IgA, has shown fluctuation during the day [124], likely due to environmental exposure. Since the serum IgA overall lacked correlation to the secretory IgA, the levels of secretory IgA in mucosal sites are poor indicators of serum IgA and vice versa. The same study reported significantly lower serum and tear IgA concentrations in young dogs (≤1 year) compared to dogs older than 1 year [124].

Variability in IgA measurements

The most established method to measure IgA levels is ELISA. However, IgA ELISA measurements are known to show variation between runs and the variability is often summarized by the coefficients of variation (CV), which is defined by the standard deviation (SD) divided by the mean. The CV is unaffected by an increase or a decrease of the mean, which usually affects the SD, and can therefore be regarded as a standardization of the SD [148]. The commonly accepted CV tolerated between measurements of the same samples is 15%.
GWAS of continuous traits

In GWAS, the studied phenotype is usually measured as a binary trait (e.g. cases versus controls) but can also be defined by continuous values. In the latter scenario, the genetic markers are associated with a deviation from the population mean. Examples of human GWAS of continuous traits are height [149], blood pressure, and hypertension [150]. In dogs, GWAS has been performed using fructosamine concentrations as a biomarker for diabetes mellitus [151]. One approach when dealing with a phenotype defined by continuous variables may be to treat it as a binary trait. This was done in the dog GWAS of serum IgE levels in LRs where levels below a certain limit (150EU) defined the case group and levels above the same limit defined the control group [152]. To define a reliable cut-off level in a study like that is the most difficult task. It must be dependent on a rigorous methodology and needs to be carefully evaluated by considering the biological effect. The use of individuals with extreme values, followed by the removal of the middle group, may be the most reliable approach to reduce the risk of the result being influenced by borderline individuals. However, the selection of a stringent cut-off level is still crucial and the drawback is the smaller sample sizes and thereby the loss of statistical power.

Paper III: Serum IgA screening of multiple dog breeds identifies breeds prone to low IgA levels

Study design and methods

We collected serum samples from 1267 dogs of 22 different breeds as part of various research projects and measured the serum IgA concentrations using ELISA (as described in Paper I). The aims were to define the normal range of serum IgA in dogs and to identify breeds with a high prevalence of low IgA concentrations. The measurements and evaluation of serum IgA were additional to the primary aims within each of the projects. The projects involved diseases represented by one or more breeds: CAD (N\text{cases}=289), Shar-Pei autoinflammatory disease (SPAID, N\text{cases}=84), Addison’s disease (N\text{cases}=58), pancreatic acinar atrophy (N\text{cases}=36), lymphocytic thyroiditis (N\text{cases}=13), diabetes mellitus (N\text{cases}=10), and steroid responsive meningitis arteritis (N\text{cases}=9) – in total 499 dogs with immune disease. As IgA production is age-dependent, 20 dogs younger than one year of age were excluded, and the final sample set consisted of 1247 dogs. The statistical calculations were performed using routine procedures and software: D’Agostino-Pearson omnibus K2 to test for normal distribution, Spearman correlation test for the effect of ‘age at sampling’ on IgA concentrations, and the Mann-Whitney test for comparisons between groups (i.e. sex, disease status, and castration.
status). The disease status was tested only within breeds and the significance level was set to p<0.05.

**Results**

The IgA concentration ranged from 0.01 to 3.0 g/l (median=0.18 g/l and mean=0.27 g/l) in the entire study population (Figure 6). The lower 95% CI was 0.26 g/l, which is similar to what has been observed in a population of crossbred dogs (0.22 g/l) [153], but higher compared to other studies in pure-breeds reporting 0.18 g/l [128] and 0.15 g/l [132, 133].

**Figure 6. IgA concentrations for all 22 breeds** plotted with a log transformation of the values and the breeds sorted based on median values. The suggested canine IgA deficiency cut-off has twice been suggested at a level of 0.15 g/l [132, 133] compared to the human well-established cut-off at 0.07 g/l.

**Eight breeds with markedly low IgA levels**

The IgA concentrations were positively correlated to the ‘age at sampling’ (p<0.0001), but no differences were seen between males and females, or castrated and intact dogs. The IgA concentrations varied widely amongst breeds reflected by the lower 95% CI of the mean that ranged from 0.05 g/l in Norwegian elkhound to 0.73 g/l in Leonberger. When applying the cut-off for human IgA deficiency (IgA<0.07 g/l), we observed 11 breeds with a high proportion (≥10%) of dogs with low IgA concentrations. The GSD (14%) and Shar-Pei (45%) were previously defined high-risk breeds for low IgA levels. The other breeds were Hovawart (32%), Norwegian elkhound (21%), Nova Scotia duck tolling retriever (NSDTR) (20%), Belgian shepherd (20%), Bullterrier (15%), American Staffordshire bullterrier (15%), GR (13%), LR (12%), and Staffordshire bullterrier (10%). However, additional studies are required to conclude if Belgian shepherd (N=10), American Staff-
fordshire terrier (N=13), and Staffordshire bull terrier (N=20) are truly high-risk breeds for low IgA levels due to the lack of information regarding ‘age at sampling’. Yet, the remaining eight breeds can be regarded as high-risk breeds for low IgA levels.

**Correlation between IgA levels and immune-mediated diseases**

In order to evaluate if the low IgA levels may be a secondary effect caused by disease, we performed correlation tests on IgA to each disease in GSD (CAD and pancreatic acinar atrophy), LR (CAD), GR (CAD), SP (SPAID), Standard poodle (Addison’s disease), Bearded collie (Addison’s disease), NSDTR (steroid responsive meningitis arteritis), Elkhound (diabetes mellitus), and Giant Schnauzer (lymphocytic thyroiditis). Interestingly, IgA levels only showed correlation to CAD (p<0.0001) and pancreatic acinar atrophy (p=0.04) in GSD, and to no other diseases in any other breed.

**Paper IV: IgA-associated loci harbor candidate genes with indicated functions in early B-cell development**

**Study design**

In this study, the aim was to identify genetic risk loci for low IgA levels in dogs, by performing GWAS in four of the breeds defined in Paper III as potential dog model breeds for IgA deficiency. GWAS was performed in 496 GSDs, 129 GRs, 128 LRs, and 94 SPs, which were the same dogs as used in Paper III except for additional GSDs. The additional GSDs were not collected in a disease study, but in a genetic study of dog behavior. All genotyped dogs had information about the ‘age at sampling’ and we used Pearson’s product moment correlation coefficient to test if any of the variables age, sex, sub-populations, or CAD were correlated to IgA levels.

**Novel GWAS approach**

In order to handle a phenotype based solely on a continuous and fluctuating variable with undefined cut-offs in GWAS, we outlined a novel approach by defining groups based on breed-specific percentiles. Moreover, by using sliding windows of different cut-offs to define the groups, we avoided a bias potentially caused by individuals on the border of the cut-off values. We then combined the results from four different GWAS. Similarly to Karlsson *et al.* [20], we defined genome-wide significance using 95% CIs, calculated from the empirical distribution of p-values observed by rerunning the GWAS with randomly permuted phenotypes 1,000 times. The permutations were performed in each percentile group (for each breed) separately, and SNPs exceeding the 97.5% upper empirical CI were defined as genome-wide significant. SNPs were considered suggestively associated if p<0.0005.
Pathway analyses
Pathway analyses were performed with the web-based program GRAIL using the PubMed text (Aug 2012) database on the associated genomic regions. Gene set enrichment testing was performed with INRICH (INterval enRICHment analysis), using 1,000,000 permutations to test the IgA associated regions for enrichments in gene sets from the Gene Ontology (GO) catalogue.

Results

German shepherds
In GSDs, we detected three genome-wide significantly associated regions. Two of these regions were defined by single SNPs. One on chromosome 23, not in LD with any other SNP, and located in the intron of GPR149 (G protein-coupled receptor 149). The other single SNP was located on chromosome 8 in high LD ($r^2>0.8$) with three SNPs, spanning ~0.5Mb. This region harbored 13 genes including SERPINA9 (Serpin Peptidase inhibitor Clade A (Alpha-1 Antiproteinase), member 9) and nine additional members of the SERPIN gene family.

The top signal was detected on chromosome 5 consisting of 14 genome-wide significantly associated SNPs in LD covering ~1Mb. When phasing the SNPs in strong LD ($r^2>0.8$, N_SNPs=17) with the top SNP covering ~1.7Mb, we identified three common haplotypes and nine rare haplotypes. Dogs heterozygous for either of the two risk haplotypes had significantly lower IgA levels compared to dogs homozygous for the common haplotype. Despite the large size of the associated haplotype it harbored only one gene: KIRREL3 (the Kin of IRRE-like protein 3).

Shar-Pei
In Shar-Pei, the top associated locus was located within the gene SLIT1 (slit homolog 1 protein). Phasing of the four linked SNPs identified two common (one risk and one control) and two rare haplotypes. Considering the common haplotypes, the IgA levels were significantly different in dogs homozygous for the risk versus the control haplotype ($p=0.0005$). Moreover, heterozygous dogs had intermediate IgA levels, significantly different compared to both groups of homozygous dogs. Next, by utilizing pre-existing whole-genome sequence data from various breeds, we detected fixation within the SLIT1 gene in GSDs. The fixed blocks spanned a 134kb region within the gene and overlapped with two of the top associated SNPs in Shar-Pei and with regions of high regulatory potential. These data indicate the causative regulatory variant(s) lie within this region, and that Shar-Pei and GSD share the same risk factor for low IgA levels. Moreover, by analyzing previously published pooled whole-genome sequencing data representing 12 wolves and 60 dogs from 14 breeds [154], we identified a 75kb domestication sweep
signal located 37kb downstream of the top associated SNP and overlapping with the fixed region in GSDs. The sweep was defined based on an increase of fixation index ($F_{ST}$) and a decrease in dog heterozygosity. A similar or more extreme differentiation is only observed in 7% of the autosomal part of the dog genome, indicating that selection of a certain variant of the SLIT1 gene may have contributed to important dog domestication features.

Pathway analyses
We identified 35 loci suggestively associated with IgA levels in the four different breeds. While the different breeds showed no overlapping association signal, we explored if common biological pathways could be found among the genes in the different risk loci. By using GRAIL, genes were connected by pathway key terms and we defined the pathways representing numerous genomic regions as the most relevant. These were ‘serum’, ‘insulin’, ‘matrix’, ‘carcinoma’, and ‘complement’. Four out of nine genes with significant GRAIL p-values were clearly involved in inflammatory responses. These genes were: the SERPINA9 and SERPINA12 genes that encode protease inhibitors, C3AR1 encoding a G protein coupled receptor for anaphylatoxin C3a (Complement Component 3a Receptor) known to activate mast cells, basophils, and eosinophils as well as smooth muscle contraction, and with implicated roles in asthma [155], and IL31 encoding Interleukin-31, which is a T-cell derived cytokine associated with chronic skin inflammation and pruritus [47]. When searching for gene set enrichment using INRICH and the Gene Ontology database, we found 51 gene sets significantly enriched in our regions. The gene sets related to transcriptional activity, hematopoiesis and platelets, cytokine responses, cell growth, and actin filament organization.

Discussion
Correlation between immune disease and IgA levels in GSDs
The samples from the dogs representing the eight breeds, defined as high-risk breeds for low IgA levels within this study, were initially collected within other immune-related disease projects. Thus, the dogs sampled were specifically selected for being cases or controls within the primary study. Therefore the incidence of disease within the sample cohort is probably not representative for the dog breed. Since the diseases were immune-related, low IgA levels could potentially influence them. Moreover, due to the reported connections between IgA levels and autoimmunity, asthma, and allergy in man [156], we had to consider the possibility that low IgA levels could be secondary to the disease. However, the IgA levels did only show correlation to CAD and pancreatic acinar atrophy in the GSD, and not to any other diseases in any other breed. This indicates that low IgA levels were not
secondary effects in the other breeds. It also suggests that there is an IgA-related CAD and an IgA-unrelated CAD, since low IgA levels clearly correlate to CAD development in GSD, but not in GR or LR. In Paper I, we detected a genome-wide significant signal associated to CAD in GSDs only when IgA levels were taken into account. Without including IgA levels as a covariate, the signal was absent. This suggests that CAD development in GSDs is influenced by low IgA levels rather than the opposite scenario: that CAD would cause low IgA levels in the dogs as a secondary effect.

Novel GWAS approach
The rationale behind the novel GWAS approach originated from the several complicating factors of the studied trait. Due to the lack of a generally accepted cut-off level to distinguish between normal IgA levels and IgA deficiency in dogs, a traditional case-control study was not appropriate. Moreover, since the IgA values differ between measurements (CV up to 15%), the use of the exact value in a continuous GWAS was not optimal. While ending up with the final option to divide dogs with different IgA levels into groups, we needed to take two additional factors into account. Firstly, since the IgA ranges vary greatly between the breeds, the division into groups needed to be breed-specific, hence the use of percentile intervals within each breed. Secondly, to avoid the strict cut-offs created between percentile intervals, we performed a series of percentile groups and added a robust case versus control analysis by way of removing dogs with intermediate values. All four GWAS were subsequently combined into one final GWAS per breed. Thus, the statistical model was developed to fit the actual phenotype.

Candidate genes

*KIRREL3 and SLIT1 in hematopoiesis*

The development of immune cells, including B-cells, is initiated in the bone marrow with the formation of blood compartments from hematopoietic stem cells in a process called **hematopoiesis**. Interestingly, in this study we found association to two genomic regions harboring candidate genes (*KIRREL3* and *SLIT1*) with implicated roles in hematopoiesis. The gene *KIRREL3* encodes a transmembrane protein widely expressed in the developing nervous system and with implicated role in synapse formation [157]. However, it is also crucial for the support and regulation of hematopoiesis in the bone marrow [158], suggesting that *KIRREL3* is important also for the formation of lymphocytes. *SLIT1* encodes a large extracellular matrix-secreted glycoprotein that functions as a ligand to the repulsive guidance receptors (Robo) family [159]. The SLIT proteins and their Robo receptors form complexes that have been demonstrated to modulate the chemoattractant-induced migration of mature leukocytes in inflammation as well as axon guidance and neural migration during development [160]. The role of SLIT1 in early hem-
atopoiesis is also demonstrated by the association between mutations in the gene with the rare and severe bone marrow failure syndrome aplastic anemia (AA) [161].

SERPINA9 in B-cell proliferation
The antigen/B-cell receptor and CD40/CD40L bindings are both important signals to initiate germinal center reactions such as CSR. This is the process in which B-cells start producing antibodies of different isotypes, followed by the differentiation of B-cells into plasma cells. SERPINA9 expression is restricted to B-cells in the germinal centers of the secondary lymphoid organs and peaks during CD40/CD40L signaling when B-cells proliferate and undergo antibody class-switching [162]. Thus, this gene is highly interesting as a candidate gene for IgA deficiency, both due to its restricted expression pattern and since it appears to play an important role in the multifaceted array of events leading to specific antibody responses.
General discussion

CAD candidate genes

The first two studies of this thesis involved the phenotype of CAD and identified gene(s) potentially involved in the etiology of this disease. The most obvious candidate gene was **PKP2** due to the location of the top associated SNPs, within and close to this gene. Moreover, one of the most associated gene in human AD is **FLG**, in which mutations affect the skin barrier. The **PKP2** gene is a highly relevant candidate gene due to its role in the desmosomes, which contribute to the strength of tissues like the epithelium and myocardium. Mutations in **PKP2** are known to cause *arrhythmogenic right ventricular cardiomyopathy* (ARVC), a disease characterized by myocardial loss and arrhythmias leading to sudden cardiac death due to defected cardiac desmosomes [163-165]. Keratinocytes express all the cardiac isoforms of desmosomal proteins, thus similar cellular responses to mutations in these tissues can be expected. Subsequently, a reduced expression of PKP2 in both keratinocytes and myocardial tissue was detected in individuals carrying ARVC-causing **PKP2** truncating mutations [143]. Interestingly, our data suggests an up-regulation of **PKP2** by the risk variants, rather than a reduced expression as in the ARVC cases. Moreover, the GSD is a breed not particularly prone to heart disease, rather the opposite [166]. Thus, the altered expression caused by our novel risk variants most likely do not affect cardiac tissues at all or possibly in a positive manner. Also, the location of the risk variants close to skin-specific enhancers indicates that the heart may not be affected. PKP2 is more expressed in the bottom layers of the epidermis and given its function in binding DPs, and thereby attaching IFs to the desmosomes, an increased expression of PKP2 may disturb the stability in the bottom layer of the epidermis. This may result in an increased allergen penetration and/or uptake by LCs in the basal layer of the skin. Structural stability of the stratum corneum is provided by the keratinocytes connected by **corneodesmosomes**, which are modified desmosomes in the stratum corneum. Abnormal degradation of the corneodesmosomes has been implicated in human skin diseases including AD [167]. Possibly, an overexpression of PKP2 in the desmosomes in the bottom layers of the epidermis may affect the differentiation of desmosomes into corneodesmosomes and thereby the stability of the stratum corneum. The connection between the **PKP2** gene
and CAD has been suggested previously in a study where it was shown that \textit{PKP2} mRNA expression was significantly up-regulated in atopic (lesional as well as non-lesional) compared to healthy skin, and that the expression of \textit{PKP2} in non-lesional atopic skin correlated with clinical severity [168]. Despite the lack of breed-matched cases and controls in this study, the results support our finding that \textit{PKP2} expression is potentially altered in CAD-affected dogs. Other functions of plakophilins include involvement in multiple signaling and metabolic processes, and also in transcriptional activity (reviewed in [169]). Thus, potentially the role of \textit{PKP2} in CAD may stretch to other functions independent from the desmosomes.

Since enhancer regions in dogs may be located far from the affected gene and with other genes in between, also the neighboring genes including, \textit{DNM1L}, \textit{YARS2}, and \textit{FGD4}, must be considered as potential candidates. The expression pattern of the four genes \textit{PKP2}, \textit{YARS2}, \textit{DNM1L}, and \textit{FGD4}, were quite different according to recorded enhancer expression from the \textit{functional annotation of mammalian genomes} (FANTOM) 5 database [170, 171]. This database provides transcription initiation activities in more than 1,000 human and mouse cell types, cell lines, and tissues, using the technique \textit{cap analysis gene expression} (CAGE). CAGE extracts, reverse-transcribes (into DNA), PCR amplifies and sequences the 5' end of the mRNA in the biological sample [172]. The fact that active promoters and enhancers are transcribed makes it possible to use CAGE to define the level of expression of specific genes in various tissues and cell types. By utilizing this database, we noted that the \textit{PKP2} gene showed the highest expression in heart, followed by the chorionic membrane cells, prostate epithelial cells, and intestinal epithelial cells. Since mutations in the \textit{PKP2} gene have been confirmed to affect the heart, it is not surprising that the \textit{PKP2} expression levels were high in this tissue. However, the high expression found in intestinal epithelial cells was more intriguing when it comes to CAD. GSDs are known to be particularly prone to gastro-intestinal problems and are also affected by CAFR. Possibly, an up-regulation of \textit{PKP2} expression in the intestinal epithelial cells may increase allergen penetrance and uptake through the intestine, by disturbing desmosome stability.

Among the other four genes located close to \textit{PKP2} the functions described for the \textit{FGD4} gene make it a possible candidate gene in addition to \textit{PKP2}. \textit{FGD4} (FYVE, RhoGEF and PH domain containing 4) is an actin filament (F-actin) binding protein with the alias name frabin. Frabins are essential during myelin formation [173], and mutations in \textit{FGD4} are known to cause neuropathies characterized by progressive muscular and sensory loss [174]. Moreover, \textit{FGD4} expression was significantly increased in DCs treated with activated House dust mite allergen (Der p 1) in both atopic and normal skin
Thus, frabins are potentially involved in the allergic reaction in skin triggered by allergens, and an increased expression of the gene may enhance an allergic response in a similar way as in the experimentally allergen-sensitized skin. Moreover, host cell actin is commonly a target for many pathogenic microbes and parasite invasion involves host cell actin reorganization in which frabins are involved. Interestingly, an inhibition of frabin proteins resulted in significantly hindered parasite invasion through the gastrointestinal tract [176]. Thus, assuming that an increased expression of frabin could result in an increased parasite invasion through the intestine, this could possibly trigger immune responses to invaded pathogens and thereby activate the allergic responses to allergens. The FGD4 involvement in DCs and neural cells and its relevance for immune cells was consistent with the FANTOM database search results showing high expression in neutrophils, monocytes, macrophages, DCs, and basophils, as well as in spinal cord, blood and thalamus.

The DNM1L gene encodes Dynamin-1-like protein, which has functions in the mitochondrial division through actin filament involvement [177] and for the maintenance and distribution of the endoplasmic reticulum [178]. DNM1L showed overall low expression across the cell types in the FANTOM database. The YARS2 gene encodes a mitochondrial protein (tyrosyl-tRNA synthetase 2) that catalyzes the attachment of tyrosine to tRNA and showed a background expression in the tested cell types in the FANTOM database. Since both DNM1L and YARS2 genes seem to have functions in basic cellular mechanisms and showed an even expression pattern across cell types, they may function as housekeeping genes. Thus, they are the less likely candidates for being the CAD-causative gene within this locus. This leaves us with PKP2 as the top candidate and FGD4, located approximately 150kb from the PKP2-locus, as an additional candidate to focus on in the follow-up studies.

IgA candidate genes

The two final studies involved IgA levels and aimed at investigating the dog as a model for IgA deficiency, by studying dog breeds prone to low IgA levels. Here, we first identified eight dog breeds predisposed to low IgA levels. We performed GWAS in four of these breeds and applied a novel approach for treating a phenotype based solely on a fluctuating continuous variable. This resulted in successful associations of, in particular, four genomic regions to the phenotype in GSDs and Shar-Pei, with the top candidate genes being KIRREL3, SLIT1, and SERPINA9. In total, 35 regions were nominally associated across all four breeds.
Both KIRREL3 and SLIT1 are genes recognized for their functions in hematopoiesis, which is highly relevant since it is known that IgA deficiency can be transferred by bone marrow transplantation [179]. Thus, the defect responsible for the deficiency seems to involve also the early stages of B-cell development. Moreover, the underlying mechanism in human IgA deficiency involves a failure of B-cells to differentiate into mature IgA-secreting plasma cells and not an absence or decrease of the B-cells themselves [180].

Considering the candidate gene KIRREL3, this gene is the only gene located within the boundaries of the associated haplotype of ~1.7Mb, which includes a 1Mb gene desert upstream of KIRREL3. Six genes are located downstream of KIRREL3, ~1.1-1.3Mb from the top SNP and ~10-180kb outside the 1.7Mb haplotype. Five of the genes do not have any obvious connection to IgA: ST3GAL4 (an enzyme involved in protein glycosylation), DCPS (an enzyme that clears eukaryotic cells of short mRNA), FOXRED1 (expressed exclusively in mitochondria), FAM118B (important for the biogenesis of small nuclear ribonucleoproteins), and SRPR (translocates proteins across the endoplasmic reticulum). The sixth gene TIRAP (toll-interleukin 1 receptor domain containing adaptor protein) is located ~100kb outside the 1.7Mb haplotype and is potentially interesting as it is involved in the signaling pathways of toll-like receptor (TLR) 2 and 4, including nuclear factor-kappa-B (NF-kB) activation [181]. NF-kB is a transcription factor initiating inflammation, and is also involved in numerous other immunological pathways including lymphatic development, B- and T-cell signaling, and immunoglobulin class-switching [182]. TIRAP polymorphisms have been associated with the systemic autoimmune disease SLE [181]. Despite the six genes located downstream from the locus, it is highly unlikely that the association would span such a large region, extending over the long-range KIRREL3 gene and 1Mb gene desert upstream of the gene, without the KIRREL3 gene being at least partly the reason for the association. However, given the great relevance of the TIRAP gene function with potential involvement in IgA class-switching it qualifies as a very good candidate gene. Possibly, causative variants located within the KIRREL3-locus affect the expression of both TIRAP and KIRREL3, and/or of any of the less likely genes (mentioned above) that are located close to the associated locus.

Considering the candidate gene SLIT1, the association in SP was limited to four SNPs located within this gene. The data showed blocks of fixation in GSDs overlapping with a potential domestication sweep within the regulatory regions of SLIT1. This indicates that this is the region harboring the causative variant(s) most likely affecting the regulation of the SLIT1 gene. However, due to the possibility that enhancers located within one gene may affect other neighboring genes, we may consider also ARHGAP19, which is the
closest additional gene to the associated locus. The encoded protein, Rho GTPase Activating Protein 19, is predominantly expressed in hematopoietic cells and is essential for the morphology and division of T-cells [183]. Due to the involvement of \textit{ARHGAP19} in lymphocytes and hematopoietic stem cells it may be a potential candidate to alter the regulation of IgA.

Ten out of 13 genes, located within the genome-wide associated locus on chromosome 8 in GSD, belong to the SERPIN gene family of Serine protease inhibitors, clade A. Serpins help control different chemical reactions by inhibiting the activity of certain enzymes. The protein centerin, encoded by \textit{SERPINA9}, inhibits trypsin-like serine proteases and may have a role in controlling the proteolytic activity surrounding the B-cells, and thereby regulating the movement and migration of B-cells while they are proliferating [184]. Thus, the relevance to the phenotype is striking and \textit{SERPINA9} qualifies as the top candidate gene within this locus.

Importantly, despite a careful evaluation of possible candidate genes based on either prospective connection between gene and phenotype and/or location near top associated SNPs, all genes in or close to the associated regions need to be considered in further fine-mapping procedures. Re-sequencing of entire loci, additional genotyping, association analyses, and functional studies will determine whether the more obvious candidate genes are involved or any of the more unlikely genes.

\section*{IgA and CAD}

\subsection*{Connection between CAD and IgA}

We detected a strong correlation between low serum IgA levels and CAD in GSDs but, despite the large sample sizes and CAD predisposition, there was no correlation between IgA levels and CAD in LRs or GRs. This indicates that there might be different biological pathways affecting the regulation of IgA levels and/or CAD in the different breeds. Comparing to human, IgA deficient patients are predisposed to allergy. However, far from all patients develop allergy and reported frequencies of allergies range from 13 to 84\% in IgA deficient individuals [185]. The great variability in frequencies between studies is likely partly due to the evaluation methods, but also due to real differences between the studied populations. Here, studies of different populations of dogs, \textit{i.e.} breeds, may be of great help. Apparently, the GSD may represent the type where low IgA is correlated with AD, whereas the LR and GR present a type where low IgA levels are unrelated to AD development. Possibly, the lower IgA levels predispose GSDs to develop CAD due to an altered defense against allergen penetrance through the skin and/or
intestine mucosa. Perhaps LRs and GRs carry protective genetic factors that shield against effects caused by lower Ig A levels, in regards to CAD. Moreover, the gene(s) involved in regulating IgA levels may possibly interact with the CAD-associated gene(s) within the *PKP2*-locus in GSDs.

**Clinical presentation of CAD in different breeds**

Different dog breeds express very similar clinical presentations of CAD and also respond comparably to treatment. This does not clearly indicate a difference in CAD etiology between breeds. However, the susceptibility to secondary infections caused by *Malassezia*, resulting in a phenotype characterized by hyperpigmentation and “elephant-skin”, is quite unique to GSDs and WHWTs. Interestingly, both breeds carry the *PKP2*-risk haplotype and the IgA status is unknown in WHWTs. Possibly, the two breeds may present with similar phenotypes due to the sharing of the *PKP2*-risk factor for CAD, in combination with low IgA levels. It would be interesting to measure the serum IgA levels in WHWTs and to genotype more CAD cases and controls of WHWTs for the *PKP2*-risk locus.

**IgA in wolves**

A population of Scandinavian wolves presented with very low serum IgA levels (median IgA=0.05 g/l) [186], whereas IgA levels in Canadian wolves presented “normal” levels (median IgA=0.18 g/l) comparable to dogs in general [187]. This may suggest that the effect of inbreeding, of both dogs and wolves (*e.g.* Scandinavian wolves, known to be inbred), can contribute to the development of immune deficiencies such as IgA deficiency. The possible reason for this may be an enrichment of causative mutations in closed populations and/or a reduced variability of immune-genes that are dependent on diversification to function efficiently. Low IgA levels likely affect the wolves’ health and fitness negatively, similarly to human and dogs, however to my knowledge AD has not been reported in wolves.
Future perspectives

While the genome structure of the dog makes it extremely useful for the initial mapping of disease loci, the presence of long regions with LD within breeds sometimes makes further genetic and functional analyses of causative variants more difficult. The work presented in this thesis represents a spectrum of what can be expected from GWAS in dogs, including long- and short-ranged associated loci harboring single or multiple genes, as well as multiple causative variants with a complicated pattern of sharing between breeds.

Since we search for novel genes involved in complex diseases, in this case CAD and IgA deficiency, we ought not only to consider the obvious candidate genes within the associated loci. Much of each gene’s function is still unknown, and what is reported for a gene may be just a small piece of all its possible functions. Gene expression in different tissues, as well as the known roles of the proteins, can give us a hint of which gene is the best candidate. However, in the follow-up studies all genes in an associated region need to be evaluated. To further prove that the actual gene gives rise to the disease through an altered expression in particular cell types or tissues, follow-up by extensive functional studies are necessary.

We have come a long way towards this goal in the CAD study. We started with phenotype classifications, sample collection, genotyping, and various genetic analyses, which have resulted in functional candidates with potential effects in four relevant cell types. The next step will be to study biological samples from dogs (carriers and non-carriers of the risk variants) to detect differences in, for example, skin pathology and mRNA expression that could affect the development of CAD. To evaluate the contribution of the risk variants to CAD development in a wider perspective, more dogs (both GSDs but also other breeds) should be genotyped. This could be done in a pilot DNA-testing project ideally prior to actions when it comes to breeding guidance. Thereby a correlation between the different variants and CAD could be assessed in a larger fraction of the GSD population. For an extension to human AD, it would be logical to sequence the PKP2-locus in samples available from human AD cases and healthy controls (or use the available human HapMap populations as controls) in order to search for variants in the regulatory regions. Then one could calculate if the allele frequencies differ be-
between cases and controls. Also, the mRNA expression of \(PKP2\) (and \(FGD4\)) could be studied in available tissue samples from human AD cases and controls.

When it comes to the IgA project, no causative variants have yet been suggested. Re-sequencing of the associated loci in dogs, with and without the risk haplotypes, would be the first step to take here. One approach to move on from these findings could also be to screen the dog candidate genes directly in human patients with IgA deficiency. The initial step could be to analyze available genome SNP and/or sequence data from patients and controls, to study the allele frequencies of variants close or within the candidate genes. Association to these genes may be restricted to families or subgroups of populations or ethnicities, which may explain the lack of association to any of these genes in previous human GWAS of IgA deficiency.

To summarize, the heterogenic human populations may hinder the findings of causative variants affecting subsets of patients groups and/or conferring a low risk, which is often the case for complex diseases. These genes may instead confer a higher risk for disease within different dog breeds and therefore be easier to detect. Thus, if we can identify disease-associated genes in dogs, these genes may explain parts of human disease etiology and thereby substantially contribute to the understanding of the human counterpart.
Populärvetenskaplig sammanfattning

Bakgrund

Vi människor är, liksom djuren, utrustade med ett välutvecklat immunförsvaret för att vi inte ska bli sjuka av det stora antal bakterier, virus, parasiter och andra mikroorganismer som finns i vår omgivning. Immunförsvaret är uppbyggt av fysiska barriärer (såsom hud och slemhinnor), ett stort antal olika immunceller och ett noggrant ordnat signalsystem. En av de viktiga komponenterna i immunsystemet är antikropparna som tillhör det anpassningsbara immunsystemet, vilket kan känna igen mycket specifika delar av proteinerna hos de invaderande mikroorganismerna.

 Ibland fungerar dock inte immunsystemet som det ska, vilket kan leda till olika immunmedierade sjukdomar. Atopisk dermatit (AD) och IgA-brist är två exempel på sjukdomar som drabbar både människa och hund, och som har att göra med ett bristfälligt immunförsvar. AD orsakas av ett överkänsligt immunförsvar som reagerar med en allergisk reaktion mot partiklar i omgivningen som egentligen är ofarliga. Kroppen aktiverar då försvarsmechanismer som respons på att IgE binder till ett specifikt ämne, s.k. allergen. IgE är en av fem olika typer av antikroppar vars normala funktion är att försvara kroppen mot parasiter. Bandningen mellan allergenet och IgE leder till en utsöndring av flera ämnen, bland annat histamin, som orsakar den allergiska reaktionen med symtom som svullnad, rodnad och vävnadsskada. IgA är den antikropp som utsöndras i allra störst mängd i slemhinnor, svettkörtlar, saliv och tårar (sekretoriskt IgA), och som även finns i blodet (serum IgA). IgA bidrar därmed till den första försvarsbarriären mot invaderande mikroorganismer och en brist på IgA är ofta kopplat till upprepade infektioner, allergier och autoimmuna sjukdomar.

 Både AD och IgA-brist orsakas av flera genetiska riskfaktorer och påverkas även av miljöfaktorer. Till exempel har studier av AD hos människa visat att förändringar (mutationer) i genen *FLG*, som kodar för proteinet Filagrin, är starkt kopplade till en ökad sjukdomsförekomst. Filagrin har en viktig funktion i att bibehålla en stark hud och mutationerna i genen orsakar en försämrad hudbarriärfunktion med ökat upptag av allergener, som leder till aktivering av immunförsvar. Även en rad andra gener som kodar för proteiner uttryckta i huden och i immunceller har kopplats samman med en ökad risk för utveckling av AD. IgA-brist har främst kopplats samman med gener som bidrar till immunförsvarets känslighet vid igenkännandet av
olika ämnen. Trots denna kunskap saknas en helhetsbild om varför de olika sjukdomarna uppkommer och en stor del av de genetiska komponenterna, som bidrar till sjukdomarna, återstår att förklara.

Många av de sjukdomar som drabbar människan förekommer även hos hunden och då ofta med olika grad av sjukdomsforemst hos olika hundraser. Detta beror på att hundarna har avlats intensivt för olika önskvärda egenskaper, vilket har bidragit till en ökad förekomst av icke önskvärda egenskaper genom en anrikning av sjukdomsframkallande mutationer. Till exempel har det rapporterats en tydlig skillnad mellan raser att drabbas av både CAD (Canine AD) och låga IgA-nivåer, vilket visar på en betydande underliggande genetisk komponent. Rasen schäfer har visats ha en ökad risk för bådadera. Andra högriskraser för CAD är bl.a. boxer, bullterrier och west highland white terrier. Låga IgA-nivåer har beskrivits hos shar-pei och beagle.

Nya forskningsresultat

I den första studien som presenteras i avhandlingen utförde vi en genetisk associationsstudie av CAD. Detta innebar att genetiska markörer spridda över hela arvensmassan undersöktes (genotyperades) hos ca 100 sjuka och 100 friska schäfrar med målet att identifiera genetiska riskfaktorer för sjukdomen. Genotyperna hos de sjuka hundarna jämfördes med de friska, vilket ledde till att vi kunde identifera en region på kromosom 27 som skiljde sig signifikant mellan grupperna och som därmed verkar vara inblandad i uppkomsten av CAD. Den starkaste kandidatgenen i denna region var PKP2, som kodar för proteinet plakophilin 2, vilket framförallt är en viktig komponent i desmosomer. Desmosomer är strukturer som sitter mellan celler och bidrar till mekanisk tålighet, särskilt betydelsefull i vävnad som utsätts för stark påfrestning som huden och hjärtmuskulaturen. Schäfrarna som ingick i studien testades också för serum IgA och vi upptäckte att låga IgA-nivåer kraftigt ökade risken att drabbas av CAD. Schäferrasen noterades även vara uppdelad p.g.a. avlsselsett och en utställningstyp av schäfer, där den senare var drabbad av CAD i högre utsträckning.

I den andra studien letade vi efter specifika genetiska förändringar i den CAD-associerade regionen i syfte att utvärdera effekten på genuttryck och därmed möjlig påverkan hos hundarna som drabbas av CAD. Vi genotypade fler hundar av olika raser vid många olika positioner inom den associerade regionen. Tack vare den utökade genotypningen och associationsanalyser, både separat i schäfer och i flera raser samtidigt, kunde vi identifiera genetiska riskvarianter vid fyra separat positioner i regionen. I ett nästa steg utvärderade vi, vid var och en av dessa fyra positionerna, vilken effekt de

I den tredje studien mätte vi IgA i serum hos ca 1300 hundar av 22 olika raser för att identifiera vilka raser som löper en ökad risk att drabbas av låga IgA-nivåer. Vi kunde se att IgA-nivåerna ökade med ålder, men att det inte fanns några könsskillnader. Medelvärdet av IgA-nivåerna skiljde sig mycket mellan olika raser. De raser där 10% eller fler av hundarna inom rasen hade nivåer under 0,07 g/l (vilket motsvarar IgA-brist hos människa) var schäfer, shar-pei, hovawart, norsk älghund, nova scotia duck tolling retriever, bullterrier, golden och labrador retriever. Eftersom vissa hundar som ingick i studien var drabbade av andra immunologiska sjukdomar studerade vi även korrelationen mellan IgA-nivåer och de olika sjukdomarna CAD, pankreasinsufficiens, Addisons sjukdom, diabetes, sköldkörtelrubbning och ”tollarsjukan”. Intressant nog såg vi endast en koppling hos schäfer där låga IgA-nivåer ökade risken både för CAD (som vi även noterade i den första studien) och för pankreasinsufficiens.

In den fjärde studien utförde vi ytterligare en genetisk associationsstudie med målet att denna gång identifiera riskfaktorer som leder till en brist av IgA. Vi valde raserna schäfer, shar-pei, labrador och golden retriever, som i den tredje studien visade sig vara högriskraser för att drabbas av låga IgA-nivåer. Vi utvecklade ett eget tillvägagångssätt för att hantera IgA-nivåer som en fenotyp och applicerade detta i associationsanalyserna, separat för varje ras. Detta ledde till att vi kunde påvisa association av totalt 35 olika genetiska regioner till fenotypen i alla fyra raserna. Dessutom kunde vi se att generna som fanns i dessa regioner var inblandade i biologiska processer, såsom inflammation, tidig immuncellsutveckling och genuttrycksreglering. De mest signifikant associerade regionerna fann vi hos schäfer och shar-pei, som också är de främst rapporterade högriskraser för låga IgA-nivåer. Hos schäfer var generna KIRREL3 och SERPINA9 de starkaste kandidatgenerna. KIRREL3 kodar för ett protein som främst är inblandat i bildandet av synapser i nervsystemet och därmed är viktig för signalering mellan celler, men som dessutom visats vara viktig vid bildandet av immunceller i bennärgen. SERPINA9 kodar för ett enzym som specifikt uttrycks i lymfkörtlar precis vid den tidpunkt då B-celler utvecklas till antikropp-producerande plasma-
celler. Hos shar-pei låg den associerade regionen mitt i genen SLIT1. SLIT-proteiner har visats vara viktiga vid den kemiskt inducerade vandringen av mogna immunceller vid inflammation samt vid nervcellers utveckling. Av intresse är att hundar i jämförelse med vargar hade reducerad genetisk variation inom den reglerande delen av SLIT1-genen, där bindningen av proteiner som reglerar genuttryck sker. Detta tyder på att hundar selekterades för någon specifik funktion som är kopplad till SLIT1-proteinen när den domesticerades från varg. När vi undersökte schäfrar mer ingående kunde vi dessutom notera att alla schäfrar saknade genetisk variation inom delar av den regulatoriska regionen av SLIT1-genen och därmed bär schäfrarna sannolikt på samma riskfaktor för IgA-brist som shar-pei. Detta kan möjligen bidra till att hela rasen schäfer är särskilt utsatt för låga IgA-nivåer.

Sammanfattningsvis har vår forskning rörande CAD och IgA-brist hos hund resulterat i identifieringen av starka kandidatgener. För CAD har möjliga genetiska förändringar föreslagits som visats ha påverkan på genuttrycket specifikt i olika celltyper. Detta ger en god grund för vidare studier, både i hund och i människa, som förhoppningsvis kan leda till en bättre förståelse om de biologiska processer som orsakar de två olika, men ändå sammanlänkande, sjukdomstillstånden. Detta kan i sin tur leda till en utveckling av nya behandlingsmetoder, anpassade efter individer med olika genetiska riskfaktor.
Acknowledgements

The studies that built this thesis were performed at the Department of Medical Biochemistry and Microbiology at Uppsala University. I would like to express my appreciation to all the people that I have worked with during my PhD studies. In particular, I would like to acknowledge the following people.

Kerstin, it has always been challenging and stimulating to work with you and I am very proud to be a part of your group. Thank you for pushing me to learn more and for your patience when things went slower than expected. Over the years you have not only taught me basic biology, and to search and connect complicated science, but also to think critically and believe in myself. Åke, you fill a special role for me and, I believe, for the whole dog group. You have a brilliant way of handling both people and dogs, and you bring calmness and stability to the group. You always listen and give valuable feedback. Göran, you introduced me to the field of genetics. I was hooked after a lecture ten (!) years ago and you cheerfully took the time to answer all my questions. This led to a master essay and article under your supervision, and I have been a part of the dog group since then. You are always positive and inspiring, and great at explaining complex biology. Gerli, thanks for sliding in as my co-supervisor and especial thanks for holding my hand during the writing process, but also for your guidance in the lab. You have great supervisor qualities by being a good listener and you remember to see the whole picture, not just the science. Cecilia, it feels like you are my fifth supervisor ☺. Thanks for helping out with all complicated paperwork, planning and for keeping an overview of all my projects, but most importantly, you always held my back.

After these many years, there has been quite a few people coming and leaving the lab. You have contributed to a stimulating environment and I have enjoyed my work thanks to all of you! Thanks to my roommates over the years: Mia, Nina, Izabella, Ingegerd, Fabiana, Brita, Maja, Lina, Maria etc. Thanks to Susanne, Ulla, Eva, and Åsa for all your help with samples and lab work. Part of the Dog group are the people at the BROAD institute and especially I want to mention Elinor, Michele and Noriko, who have been around as long as I have ☻. Elinor, you picked up a new and quite lost student at Boston airport quite a few years ago. Thank you for introducing me to Boston and the Broad institute, and for making me feel welcome. You are encouraging and one of the most skillful scientists I have ever met.
I have also enjoyed the contact with all the breeders, dog owners, veterinarians, and technicians, who made this research possible.

To my clever co-authors of the papers: it has been, and still is, a pleasure working with you. Special notions to: Mia, we complement each other extremely well, you write and I make the tables… I suggest we always have at least one manuscript going! Marcin, thanks for your patience when teaching me R, you are a great teacher but also a very skilled scientist. I enjoy our scientific discussions! Kerstin B, you have given me the opportunity to see the reality and not just the samples or data points. I am impressed by your work and knowledge. The way that you treat your patients and their owners is outstanding! Sergey, I am happy you joined the Atopy project, your contributions to the project are invaluable. I enjoy working and discussing science with you. Fabiana, I will miss you as my roommate and colleague. You are very talented both in the lab and with computer work, and you always come with such good ideas.

Ett extra tack till mina kollegor och vänner: Mia och Nina, vad vi har varit med om mycket tillsammans. Ni är sådana där vänner man alltid har nära hjärtat trots att man inte hinner ses så ofta. Mia, speciellt jättetack för att du läst och kommenterat min kappa (och de flesta andra skrifter) ett antal gånger!


Vänner, grannar och släktingar, tack för att ni ger mig en annan vinkel på livet 😊.
References


40. Ameratunga R, Woon ST, Gillis D, Koopmans W, Steele R. New diagnostic criteria for common variable immune deficiency (CVID), which may assist with decisions to treat with intravenous or subcutaneous immunoglobulin.


66. Saevik BK, Ulstein TL, Larsen HJ. Evaluation of a commercially available enzyme-linked immunosorbent assay for the detection of allergen-specific IgE


120. Hammarstrom L, Carlsson B, Smith CI, Wallin J, Wieslander L. Detection of IgA heavy chain constant region genes in IgA deficient donors: evidence


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1128

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)

Distribution: publications uu se
urn:nbn:se:uu:diva-259606