UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1328

Genetic Studies of Immunological Diseases in Dogs and Humans

MATTEO BIANCHI

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2017

ISSN 1651-6206
ISBN 978-91-554-9901-3
urn:nbn:se:uu:diva-319962

Dissertation presented at Uppsala University to be publicly examined in B41, BMC, Husargatan 3, Uppsala, Monday, 5 June 2017 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Marta Alarcón-Riquelme (Center for Genomics and Oncological Research (GENYO), Granada, Spain).

Abstract

Bianchi, M. 2017. Genetic Studies of Immunological Diseases in Dogs and Humans. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1328. 68 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9901-3.

This thesis presents genetic studies aiming at enlarging our knowledge regarding the genetic factors underlying two immune-mediated diseases, hypothyroidism and autoimmune Addison's disease (AAD), in dogs and humans, respectively.

Genetic and environmental factors are indicated to contribute to canine hypothyroidism, which can be considered a model for human Hashimoto's thyroiditis (HT). In Paper I we performed the first genome-wide association (GWA) study of this disease in three high-risk dog breeds (Gordon Setter, Hovawart and Rhodesian Ridgeback). Using an integrated GWA and meta-analysis strategy, we identified a novel hypothyroidism risk haplotype located on chromosome 12 being shared by the three breeds. The identified haplotype, harboring three genes previously not associated with hypothyroidism, is independent of the dog leukocyte antigen region and significantly enriched across the affected dogs. In Paper II we performed a GWA study in another high-risk breed (Giant Schnauzer) and detected an associated locus located on chromosome 11 and conferring protection to hypothyroidism. After whole genome resequencing of a subset of samples with key haplotypes, we fine mapped the region of association that was subsequently screened for the presence of structural variants. We detected a putative copy number variant overlapping with the upstream region of the *IFNA7* gene, which is located in a region of high genomic complexity. Remarkably, perturbed activities of type I Interferons have been extensively associated with HT and general autoimmunity.

In Paper III we performed the first large-scale genetic study of human AAD, a rare autoimmune disorder characterized by dysfunction and ultimately destruction of the adrenal cortex. We resequenced 1853 immune-related genes comprising of their coding sequences, untranslated regions, as well as conserved intronic and intergenic regions in extensively characterized AAD patients and control samples, all collected in Sweden. We identified *BACH2* gene as a novel risk locus associated with AAD, and we showed its independent association with isolated AAD. In addition, we confirmed the previously established AAD association with the human leukocyte antigen complex.

The results of these studies will hopefully help increasing the understanding of such diseases in dogs and humans, eventually promoting their well-being.

Keywords: complex disease, immunogenetics, autoimmunity, GWAS, NGS, canine model, dog, hypothyroidism, Addison's disease, IFNA, BACH2

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

© Matteo Bianchi 2017

ISSN 1651-6206
ISBN 978-91-554-9901-3
urn:nbn:se:uu:diva-319962 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-319962)

"*Non quia difficilia sunt non audemus, sed quia non audemus difficilia sunt*"

"It is not because things are difficult that we do not dare, it is because we do not dare that things are difficult"

Lucius Annaeus Seneca

To my family

List of Papers

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

I **Bianchi, M.**[*], Dahlgren, S.[*], Massey, J., Dietschi, E., Kierczak, M., Lund-Ziener, M., Sundberg, K., Thoresen, S.I., Kämpe, O., Andersson, G., Ollier, W.E.R., Hedhammar, Å., Leeb, T., Lindblad-Toh, K., Kennedy, L.J.[#], Lingaas, F.[#], Rosengren Pielberg, G.[#] (2015) Multi-Breed Genome-Wide Association Analysis for Canine Hypothyroidism Identifies a Shared Major Risk Locus on CFA12. *PLoS One*, 10(8):e0134720 [*]These authors contributed equally to this work [#]These authors contributed equally to this work

II **Bianchi, M.**, Sundberg, K., Rafati, N., Karlsson, Å., Andersson, G., Kämpe, O., Hedhammar, Å., Lindblad-Toh, K., Rosengren Pielberg, G. The Type I Interferon Gene Cluster is Associated with Hypothyroidism in a Swedish Giant Schnauzer Dog Population. *Manuscript*

III Eriksson, D.[*], **Bianchi, M.**[*], Landegren, N., Nordin, J., Dalin, F., Mathioudaki, A., Eriksson, G.N., Hultin-Rosenberg, L., Dahlqvist, J., Zetterqvist, H., Karlsson, Å., Hallgren, Å., Farias, F.H.G., Murén, E., Ahlgren, K.M., Lobell, A., Andersson, G., Tandre, K., Dahlqvist, S.R., Söderkvist, P., Rönnblom, L., Hulting, A.-L., Wahlberg, J., Ekwall, O., Dahlqvist, P., Meadows, J.R.S., Bensing, S., Lindblad-Toh, K., Kämpe, O.[#], Rosengren Pielberg, G.[#] (2016) Extended exome sequencing identifies *BACH2* as a novel major risk locus for Addison's disease. *Journal of Internal Medicine*, 280(6):595–608 [*]These authors contributed equally to this work [#]These authors contributed equally to this work

Reprints were made with permission from the respective publishers.

Related works by the Author
(Not included in this thesis)

I Kierczak, M., Jablonska, J., Forsberg, S., **Bianchi, M.**, Tengvall, K., Pettersson, M., Scholz, V., Meadows, J.R., Jern, P., Carlborg, Ö., Lindblad-Toh, K. (2015) cgmisc: enhanced genome-wide association analyses and visualization. *Bioinformatics*, 31(23):3830–1

Contents

Abbreviations

21-OH	21-hydroxylase
AAD	Autoimmune Addison's disease
AITD	Autoimmune thyroid disease
APC	Antigen presenting cell
APS	Autoimmune polyendocrine syndrome
bp	base pair(s)
CFA	*Canis familiaris* autosome
CNV	Copy number variant
DLA	Dog leukocyte antigen
DNA	Deoxyribonucleic acid
GD	Graves' disease
GS	Gordon Setter
GWA	Genome-wide association
HC	High coverage
HLA	Human leukocyte antigen
HT	Hashimoto's thyroiditis
HV	Hovawart
IBD	Identity by descent
IFN	Interferon
Kb	Kilobase pairs
LC	Low coverage
LD	Linkage disequilibrium
MAF	Minor allele frequency
Mb	Megabase pairs
NGS	Next generation sequencing
OR	Odds ratio
QC	Quality control
RNA	Ribonucleic acid
RR	Rhodesian Ridgeback
SAR	Swedish Addison Registry
SNP	Single nucleotide polymorphism
T3	Triiodothyronine
T4	Thyroxine
Tg	Thyroglobulin
TgAA	Thyroglobulin autoantibody
TPO	Thyroid peroxidase

TSH	Thyroid-stimulating hormone
TSHR	Thyroid-stimulating hormone receptor
UTR	Untranslated region
WES	Whole-exome sequencing
WGS	Whole-genome sequencing

Introduction

For a very long time geneticists have tried to answer the question regarding how the genetic makeup of an individual, the genotype, links to specific characteristics shown by that individual, the phenotype. In a simplistic situation, when a single gene controls the phenotype, the trait displayed by the individual is generally explained by the genetic variation of a single locus. Conversely, a considerable challenge has been the identification of the wide spectrum of genetic determinants underlying complex traits, and especially human complex diseases. In this class of diseases, a multifactorial genetic etiology contributes to the individual susceptibility. Unraveling and understanding the genetic architecture underlying the disease risk would mean moving forward towards improved disease diagnostics, treatment and prevention, thus eventually realizing the so called personalized medicine concept.

Although there have been significant efforts to decipher the complex disease genetics puzzle, there are still missing pieces that need to be added to the framework. In this context, the domestic dog has provided great help in trying to achieve this objective. Dogs spontaneously develop a wide spectrum of diseases shared with humans and it is likely that the same biological pathways and mechanisms are involved in disease susceptibility in both species. Moreover, the domestic dog has a genomic structure highly amenable to the discovery of genetic disease risk loci, thus facilitating the identification and the placement of novel pieces in the puzzle. Nevertheless, the genetic knowledge gained by studying the domestic dog might also contribute to the wellbeing of this species through the implementation of specific diagnostic tools and suitable breeding strategies.

Among the several diseases affecting dogs and humans, immune-mediated disorders still represent an intricate class of pathologies. They develop in genetically susceptible individuals and involve a complex interaction between the immune system and environmental factors. Autoimmune diseases represent a class of pathologies in which the immune system mounts a response against self-molecules, eventually causing the impairment of the target tissue or organ. Autoimmune diseases affect approximately 8% of the western-world population; similarly, several dog breeds are characterized by a high prevalence of these common disorders. Significant advances in molecular techniques and sequencing technologies have in recent years provided an innovative impulse to study immune-mediated and autoimmune

diseases and to identify novel susceptibility loci underlying their development [1-8]. Genome-wide association and immune-specific fine mapping studies, as well as large-scale meta-analyses, have identified a great number of disease-associated single nucleotide polymorphisms in humans. Next-generation sequencing has revolutionized the search of disease-associated variants allowing scientists to resequence whole genomes or selected portions of them at an extremely high speed and with constantly reduced costs. Furthermore, this cutting-edge methodology allows the detection of rare and population-specific variation, which is likely to have an important contribution in disease etiology in humans. Similarly, even though a number of disease-susceptibility loci have been identified in dogs using well-established genome-wide association approaches, canine geneticists are progressively shifting to next generation sequencing in order to complement, and eventually replace, current mapping methodologies. Despite the great advancements in next generation sequencing technologies and applications in the majority of both canine and human immune-mediated diseases, a significant proportion of the disease-predisposing genetic risk has not been unraveled yet. Moreover, an additional challenge will be to understand the function of the identified genetic factors, to reveal how they participate in the underlying disease biology.

Starting from the above-mentioned aspects, this thesis presents the genetic studies that allowed us to identify novel candidate loci associated with two immune-mediated diseases in dogs and humans, represented by canine hypothyroidism and human Addison's disease, respectively. By employing genome-wide association and next generation sequencing methodologies, we sought to enlarge our modest knowledge regarding the genetic determinants underlying these diseases. Filling the knowledge gaps regarding the genetics of hypothyroidism using the dog as a model is of major interest due to the prevalence and the impact of this disease in both dogs and humans. Similarly, the discovery of novel genes increasing the susceptibility to human Addison's disease can represent a step forward towards the realization of personalized medicine.

Genetic variation and heritable traits

Apart from a few exceptions, every cell of the body contains deoxyribonucleic acid (DNA), which bears the information necessary for the organism development, growth, functions and reproduction. Genetic variation derives from changes (*i.e.* mutations) in the string of nucleotides that constitutes the DNA. Mutations, or genetic variants, can involve single or multiple nucleotides and they can spontaneously arise at different genome locations. The simplest and most frequent variants are single nucleotide substitutions, known as single nucleotide polymorphisms (SNPs). There are more than 100

million validated SNPs cataloged in dbSNP 150, and approximately 3.5 million SNPs are harbored in an individual genome [9]. According to their location in the genome, SNPs can be classified as intergenic and genic, which is typically further split into coding (synonymous, missense, nonsense), and non-coding (intronic and untranslated regions).

Differently from SNPs, structural variation derives from changes involving more than one nucleotide. Overall, these genetic variants range from short insertions/deletions (indels) to large chromosomal rearrangements [10, 11]. Structural variants represent an important source of genetic variability and have the ability to reshape the gene/genomic landscape, thus significantly contributing to the phenotypic variation.

Genetic variants can be either neutral, when having no effect on the reproductive success of the individual, or functional. Once a variant appears, its fate is dictated by genetic drift and selection. Functional genetic variation is the substrate upon which selection acts, and selection in turn represents evolution's engine. On one hand, this variation represents the necessary source of diversity that is fundamental to increase the reproductive success of individuals, especially in response to environmental modifications; on the other hand, it represents the source of detrimental changes that have a negative impact on the phenotype and thus on the fitness of the individual. Simplistically, beneficial variants raise their frequency and eventually reach fixation through positive selection, whereas detrimental variants reduce their frequency and are eventually removed through purifying selection, thus resulting in a decrease of genetic variation in both cases. Alternatively, different alleles at a locus are maintained through balancing selection, therefore resulting in increased levels of heterozygosity. In general, the specific variants effect depends on environmental conditions, and a certain effect might derive from the collective contribution of all the variants occurring in a genetic locus. Genetic variants can occur in somatic cells, and in this case only their clones will carry such genetic differences. However, if genetic variants arise in germ-line cells, they might be transferred to the next generation, and the linked traits will be heritable.

At this point, an important question is whether the phenotypic variation we observe is entirely due to genetic factors or whether environmental factors contribute to it. In simple terms, this refers to the extent of the resemblance between parents and offspring [12]. Heritability is formally defined as the ratio between genotypic and phenotypic variance in a population, and it describes the contribution of genetic factors to the phenotypic variation. The phenotypic variance can be partitioned into unobserved genotypic and environmental factors, and in this case the above-mentioned ratio defines the broad sense heritability (H^2). Moreover, the genotypic variance can be partitioned into the additive, dominant and epistatic genetic effects, thus defining the narrow sense heritability (h^2) if only the additive genetic effect is considered in the ratio [13]. A high heritability reflects that an individual pheno-

type can very well predict the genotype in a population with certain trait prevalence. Consequently, the probability to detect a gene with large effect increases with higher heritability, although this estimation does not inform about the trait genetic architecture [13].

Complex diseases in humans

Among the plethora of heritable traits, disease phenotypes have represented a compelling target for genetic research due to their substantial impact on individuals' life and global health. Mendelian diseases are the simplest class of genetic diseases and are usually caused by highly penetrant and pathogenic deleterious mutations in a single gene (*i.e.* monogenic diseases), and most often the mutation is located in an exon. In general, these detrimental mutations compromise reproductive success and are selectively purged by purifying selection, even though a mutation-selection balance usually permits their maintenance at low frequencies in a population [14-16]. Monogenic disorders tend to segregate in families according to the traditional Mendelian inheritance patterns, but phenotypic heterogeneity and incomplete or age-dependent penetrance might complicate the analysis of disease pedigrees [17]. These mechanisms might blur the distinction between monogenic and complex diseases [18, 19].

Complex diseases are generally common clinical conditions determined by multiple genetic loci, as well as environmental factors. These disorders typically aggregate in families, but their segregation is not consistent with classical Mendelian inheritance, thus confirming the hypothesis of a polygenic etiology. A wealth of common SNPs (minor allele frequency MAF > 1%) has been associated to complex diseases; apart from a few examples of variants having a large impact [20], most of them have a modest effect on disease susceptibility and overlap regions with a potential regulatory function. From an evolutionary point of view, variants with small effect are subjected to a less pervasive purifying selection and could be maintained in the population at moderate frequencies. Moreover, it has been shown that genomic regions overlapping these variants are in some cases enriched in signals of positive selection. This could suggest a role of regulatory variants in the adaptation to dynamic environmental pressures throughout human history and evolution [21, 22]. In complex diseases, the associated common SNPs with modest effect contribute only a very small fraction to the total disease heritability. Moreover, a large proportion of the disease genetic variance still needs to be explained, even though noteworthy efforts have been recently made in this direction. A significant number of variants, yet to be determined, are likely to contribute to the overall disease genetic variance, raising the so-called "missing heritability" issue [23]. The "missing heritability" concept and the resulting assumption on the genetic architecture of complex

diseases can be described either by the dated common disease – common variant (CDCV) hypothesis or by the more recent common disease – rare variant (CDRV) model. In addition, different complex diseases (for instance psychiatric and immune-mediated complex diseases) might be characterized by slightly different genetic architectures. However, it is likely that both common and rare variants contribute to complex diseases. Subsequently, a large number of common variants with small effect complemented by rare variants with larger effect could be involved in the disease genetic etiology together with interacting epistatic and epigenetic effects [24-26].

The dog, man's best friend in the study of complex diseases

The domestic dog is considered man's best friend for its way of being a unique companion animal. It is an invaluable friend not only in our daily life, but also in the long-term challenge of gaining knowledge about human disease. The domestic dog has been proven to be an effective animal model to tackle the considerable challenge of understanding the genetics underlying complex diseases that are difficult to unravel in humans [27-29].

Dogs have undergone a strong and persistent artificial selection throughout their history. This has led to the enrichment of particular alleles in the population, thus resulting in an incredibly large phenotypic diversity and in the subsequent creation of the hundreds of different breeds existing today [30-33], which represent homogenous isolated populations with decreased genetic heterogeneity [34]. Many modern dog breeds have derived from a recent bottleneck, during which a few valuable individuals (popular sires) were selected as founders of the different breeds. Modern dog breeds show an enrichment of the founder genetic character that has been continuously selected in the population [35, 36]. Breed creation has occurred within the past few hundred years, thus not allowing recombination to break the selected genomic loci. Recombination is a paramount event occurring during meiosis, which underlies sexual reproduction and the transmission of genetic information over generations. Recombination breaks down haplotypes, in other words stretches of DNA containing loci that are inherited together, thus resulting in the decay of linkage disequilibrium (LD). The continuing strong selection has kept allele frequencies skewed at multiple loci and has maintained LD significantly high within each dog breed [30, 31].

Before the creation of breeds, the dog underwent a primary bottleneck during its domestication from grey wolf, which has been estimated to have occurred about 5,000-30,000 years before present [31, 37]. It has been hypothesized that domestication may have resulted from artificial selection of wolf pups based on traits important for hunting or guarding. Alternatively,

wolves may have undergone domestication themselves when they started to scavenge close to human settlements during the transition from a nomadic to a stationary lifestyle in the context of agricultural revolution [37, 38]. During the period of modern breed creation, the pre-breed domesticated dog population was characterized by low LD due to the long time-span since early domestication that had allowed the generation of short haplotypes. The two bottlenecks throughout dog demographic history have substantially shaped the canine genome structure, which maintains the signatures of these two events: extensive LD within breeds resulting in long haplotypes, low LD level across different breeds resulting in short haplotypes. Compared to humans, the dog is characterized by a genomic landscape remarkably structured and highly amenable to the discovery of causal genetic loci. Humans show much lower LD than individual modern purebred dog breeds due to higher genetic admixture and prevalence of natural selection, which is responsible for the slow increase in frequency of alleles conveying greater fitness in specific environments [39, 40]. Artificial selection, which may not imply any advantage in terms of reproductive success, has been strengthened in modern dogs and resulted in a quick and extensive selection of genomic loci and increased levels of homozygosity precluding the detection of recombination events [30, 31]. The strong artificial selection underpinning of the breeding programs has additionally caused the overrepresentation of certain diseases in different breeds. This reflects either enrichment of deleterious alleles during the recent breed creation bottleneck, association of detrimental variants that hitchhiked with the loci underlying the selection phenotypic target, or existence of pleiotropic effects of the selected variants that control both a selected morphologic phenotype and a potential disease phenotype [30, 41, 42].

Dogs naturally develop the similar diseases as humans do, which make them excellent spontaneous animal models for investigating several human pathologies. They share a wide range of autoimmune, endocrine and cardiovascular diseases, cancers and nervous system related diseases [34, 42]. Furthermore, disease clinical progression and symptoms are often highly comparable between the two species. In some western countries, dogs also have really accurate and extensive family and clinical records, which make the characterization of the sample cohort extremely precise and reliable. Moreover, pet dogs nowadays share the same environment as humans do; this is useful when considering the influence of environmental factors on complex disease etiology. Exerting the potential of the domestic dog in disentangling the genetics of complex diseases means not only primarily promoting the health of this species, but also providing new and alternative discovery perspectives with regard to human disease genetics.

Mapping complex diseases: principles and challenges

Examples of disease mapping in the early times

Linkage mapping was generally based on simple sequence length polymorphisms and it has been remarkably successful for identifying genes associated to Mendelian diseases both in humans [43, 44], despite the issue of accessing genetic material from large multigenerational families, and in dogs, in which very large pedigrees are instead available and directly exploitable [45, 46]. However, in human complex disease genetic research this approach did not prove particularly fruitful mainly due to a major complicating factor such as the polygenic nature of these disorders, as well as the difficulties in detecting small effect sizes of common genetic variants and the issues in defining precise phenotypes [47-49]. Similarly, the study of complex diseases in dogs has a limited effectiveness using such methodology. Differently from linkage analysis, candidate gene studies focus on a particular gene or region that is selected based on a priori knowledge about its function and its putative role in the complex disease etiology. Genetic variation in the selected gene or region is subsequently evaluated for association with the disease. This approach has drawn criticism due to several factors (choice of candidate loci, statistical power, confounders) that might influence the study outcomes and therefore prevent replication [50]. Despite the difficulties, these methodologies were able to reliably identify a few common variants associated with complex diseases, but these findings were restricted to loci with large effect [51-55].

Genome-wide association studies

Genome-wide association (GWA) mapping has provided an innovative and effective strategy for the study of complex disease genetics, due to the possibility to screen the whole genome based on common genetic variation in large sample sets. The basic idea of this approach is to take advantage of the LD existing between a disease locus and one or more genetic markers that are therefore informative of the whole haplotypes on which they are located. This enables one to find a marker-disease association and to pinpoint a genomic locus carrying the mutation that contributes to the susceptibility to the disease. For this purpose, the most commonly used genetic markers are SNPs. SNPs are codominant and generally biallelic markers. The HapMap project in humans and the different genome sequencing projects in different species, including the domestic dog, have provided a large spectrum of these molecular markers with known chromosomal location that can be effectively used in GWA analyses. SNPs provide a higher coverage throughout the genome and their genotyping could also be automated, offering the possibility of developing high-throughput technical platforms, which dramatically in-

creases the productive capacity. Joint efforts between public, as well as private institutions and companies, have led to the development of SNPs chips used in genomic research [31, 41, 56-58]. In a classical case control design, GWA analysis seeks to identify differences in allele frequencies between cases showing the disease and controls not showing it. Using unrelated individuals and a set of common SNPs reasonably spread across the genome depending on its LD degree, a SNP is considered associated with the disease if the frequency of an allele is statistically significantly higher/lower in one of the phenotypic group compared to the other [47, 59].

In order to avoid false positive results, the statistical significance threshold rejecting the null hypothesis ($\alpha = 0.05$) has to appropriately take the number of independent tests into account. If not properly handled, a major confounding factor that may also affect GWA analysis is stratification. On one hand population stratification derives from the presence of distinct populations or genetic ancestries within the same cohort. On the other hand, additional sources of stratification might be represented by any qualitative attribute that strongly correlate with the case and control phenotypes. In this scenario, a positive association could be due to a difference in allele frequencies capturing the correlation with the confounders, rather than reflecting a difference truly associated with the disease [60]. An additional confounder is the possible presence of cryptic relatedness, which may be relevant in those animal models, especially the domestic dog, characterized by the extensive use of popular sires [61]. Cryptic relatedness, as stratification, structures the samples set and it should therefore be properly accounted for. A statistical parameter called inflation factor (λ) assesses if a GWA study shows stratification and measures its degree in the analyzed samples [62].

In order to face these issues and avoid inflation, different strategies have been described [60]. These include mixed model based methods, which are generic and adaptable approaches in order to correct for the occurrence of structured data in GWA analyses [63-65]. A mixed linear model is a statistical extension of a simple linear model, which is defined by the formula:

$$y = Xb + e$$

where y is the individual phenotype, X is a matrix including the genotypes at each SNP (fixed effect) tested for association with the phenotype, b is the unknown variable (regression coefficient) representing a vector that defines the strength of association between the SNP and the phenotype and e represents the residuals (difference between the real and the predicted phenotype value). The extended version of a simple linear model (*i.e.* mixed model) can include factors that are random variables and might influence the phenotype (random effects). A mixed model is described as:

$$y = Xb + Zu + e$$

where the new parameters Z and u represent a matrix for the random effect (for instance the genomic kinship matrix) and a vector containing the random effects, respectively [66].

Furthermore, the success of a GWA study depends on the number of the segregating markers used in the analysis. Since GWA analysis is dependent on at least one segregating SNP tagging the causative haplotype, the amount of SNPs required to scan the genome depends on the level of LD in the study population. For instance, a higher number of markers (millions of markers) are required to screen the human genome that is characterized by a low LD degree, whereas a lower number of markers (thousands of markers) are sufficient to scan genomes with higher LD level (*i.e.* dog breeds). Besides a reliable phenotype, the correct cohort sample size is critical for a successful GWA study, in terms of power for detecting large or small effects of specific variants associated with the disease [30, 31, 62]. In humans, the modest effect sizes of the variants associated to certain complex diseases have made extremely large sample sets (~10,000-40,000 or more cases and controls) [67] necessary in order to unveil such small genetic contributions [23]. In this context, the domestic dog could be of great help. The reduced genetic heterogeneity and the high prevalence of specific diseases in different dog breeds suggest an overall presence of a few loci of greater effects underlying each disease [30, 31, 34], and subsequently the need of a lower number of individuals compared to human studies. In an extreme situation, extensive homozygous regions could be fixed in most dogs within a breed as a consequence of significantly strong artificial selection on a desirable phenotype predisposing for a disease. In this scenario, a strategy based on homozygosity mapping may be applied in order to identify the disease-associated locus [68, 69].

Meta-analysis

The implementation and application of meta-analysis has been especially compelling in human genetics, in which only a small proportion of the heritable component for complex diseases has been identified due to the small effect sizes of the GWA studies associated variants [23, 70]. GWA studies meta-analysis is able to boost the detection of association signals by increasing the sample size. This is obtained by pooling the results from each GWA study, thus statistically synthesizing the genetic information coming from multiple standalone analyses. In addition to several preliminary planning, organizational and operational criteria, the analytical approach to be used in the meta-analysis should be chosen with attention [71, 72]. According to a current and widely used approach, different studies are given different weight according to their precision, which depends on their standard error [70]. Meta-analysis could be conducted by applying a fixed effect model (it is assumed that a common genetic effect lies beneath every single GWA

study) or a random effect model (it is assumed that the single GWA studies are assessing different effects). In meta-analysis, an important confounding factor is represented by statistical heterogeneity, which defines the diversity of the different studies in terms of experimental designs, protocols, analyses and outcome. Theoretically, a fixed effect model should be used in absence of heterogeneity, whereas a random effect model in presence of diversity [70-72]. The estimation of heterogeneity is hard to evaluate in presence of a few single studies and its estimation could possibly be invalid for all the studies combined together and for all the variants' effects. The fixed effect model is nowadays a popular approach for meta-analysis due to its notable discovery power [70, 72].

Next Generation Sequencing

Next generation sequencing (NGS), alternatively called massively parallel sequencing, is the cutting-edge biotechnology that has represented a revolution in the genomic field. It has allowed scientists to decode the nucleotides included in a genome at an extremely high speed and with significantly reduced costs and increased efficiency compared to Sanger sequencing, which had dominated the field for a long time. If we compare the overall cost of the Human Genome Project about 15 years ago (~US$3 billion) to the cost currently needed to resequence an entire genome (US$1,500-5,000), we can immediately understand the impact that this technology has had and is having on genetic research [73-76]. On top of that, it has been really recently claimed that resequencing a whole human genome will cost US$100 in the upcoming years.

With NGS, thousands to millions of sequencing reactions are performed in a massively parallel fashion, thus enabling extraordinary process scalability. With this technology, genomic DNA is sheared into smaller fragments that undergo ligation to specific adapters, creating sequencing libraries. In principle, the libraries are then incorporated in the sequencing instrumental core and randomly and digitally deciphered while subjected to sequencing [77, 78]. The generated sequence reads are subsequently aligned to a reference genome in order to identify similarity or variable sites. Most importantly, while enabling a robust genome-wide single nucleotide variant calling, NGS allows a comprehensive assessment of low frequency and rare genetic variants, which are likely to have an important genetic contribution in complex disease etiology. Rare variant association could be evaluated by aggregating rare variants in genes or different biological units. Subsequently, these defined biological intervals could be used as the units of association. Aggregate analyses could be performed by employing burden test (assumption of uniform variant effects) or variance-component test (assumption of different variant effects).

One of the most widely used and economically accessible NGS platforms is the Illumina sequencing technology that generally produces short paired-end reads (100-200 bp) [76, 77]. In spite of being high-throughput and providing accurate base level information, short reads sequencing technology might struggle to identify and resolve the wide range of structural variants and repetitive regions spread in the genome. New technologies have recently been developed with the goal of increasing sequencing read lengths (> 10 Kb) [3, 79]. Longer reads are expected to significantly facilitate *de novo* assemblies, to greatly contribute to the attempt to fill the gaps in the current reference genomes, to robustly identify structural rearrangements and to help with the detection of variation in repetitive genomic regions. These platforms will surely represent a great prospect in the future of genomic research, especially because it has been hypothesized that structural variation might account for a fraction of the genetic variance underlying complex diseases.

The potential of NGS is enormous, "with one's imagination being the primary limitation to its use" [80]. However, more and more high-standard computational infrastructures and bioinformatic capabilities are currently necessary to promptly decode the huge amounts of data that is continuously produced. On top of that, this enduring and overwhelming load of data requires huge storage capacities, which probably nowadays represents the real NGS bottleneck.

Targeted resequencing

Through targeted resequencing, selected genomic regions are captured and enriched from a DNA sample prior sequencing [81]. These genomic regions could be comprised of the whole exome, a subset of genes and loci involved in specific diseases or pathways, or a chromosome of interest [82]. A typical strategy to capture and enrich selected genomic regions is represented by in-solution sequence capture technology [83]. The individual sequencing libraries are hybridized to specific biotinylated probes in order to capture the target fragments. After hybridization to the target regions, the probes are separated from the un-hybridized fragments through the interaction with modified magnetic beads covered with streptavidin [81, 83, 84]. With respect to resequencing of whole genomes, targeted resequencing provides an increased coverage (how many times a nucleotide is sequenced) for single genome position and allows processing more samples at a much lower cost.

From association to function

Human GWA studies have yielded a myriad of common variants strongly associated with complex diseases. However, it is likely that these variants are in LD with the disease causal variant, rather than having a biological

function themselves [26]. Fine mapping seeks to narrow down the GWA analysis associated regions and the number of candidates by detecting and genotyping a set of denser variants, as well as subsequently performing prioritization according to the variants functional potential. In the domestic dog, the genome structure of this species (high LD within breeds, low LD across breeds) motivates an alternative approach. After the detection of a wide LD interval associated in one breed, fine mapping could be performed by employing another breed that is affected by the same disease and carries the same identical by descent (IBD) causative variant(s) contributing to the disease. Nevertheless, the resulting shorter candidate region is subsequently screened for variants functional prioritization.

In order to fine map a region of association, a wider and much denser spectrum of variants present in the associated LD interval needs to be genotyped or imputed with high confidence. Imputation predicts genotypes not assayed in a target group of individuals by using reference panels, often represented by large-scale resequencing datasets, that capture as much haplotypic genetic variation as possible [85]. In humans, imputation of variants by using population-specific reference panels has resulted in high quality and accurate results, especially for rare variants genotype prediction. Similarly to the large-scale resequencing projects that have been taken up and eventually provided a wealth of information about human genetic variation [86-88], an international consortium aiming at resequencing 10,000 dog genomes has been recently initiated (http://www.dog10kgenomes.org/).

In humans, an alternative fine mapping approach is to employ custom genotyping arrays collaboratively designed by international consortia. These arrays are based on data generated from robust GWA and candidate gene studies, as well as resequencing initiatives, and they include a selection of common and rare variants implicated in specific complex diseases [89-91]. For instance, the Immunochip contains ~200,000 variants that serve to perform deep replication of major immune-mediated diseases and fine mapping of loci robustly associated with these disorders in GWA studies [89, 90].

The majority of human GWA studies associated variants are located in introns and in genomic regions with a strong regulatory potential, such as promoters, enhancers and silencers, thus corroborating the notion that non-coding regions might have important biological functions [92-94]. This also emphasizes that gene regulation is likely to be a major player underlying the etiology of complex diseases. For instance, gene regulation is often tissue-specific and might operate at any time point during an individual's life. Moreover, gene regulation might vary between diverse cell types uniquely reacting to different environmental stimuli. Evaluating the overlap with conserved elements and the enrichment of GWA studies associated variants in diverse functional annotated classes (DNase I hypersensitivity sites, histone marks, transcription factor binding sites) might provide insights about their biological role. ENCODE, NIH Roadmap Epigenomics and Fantom5 are

projects aiming at supporting functional annotations of regulatory variants [92, 93, 95]. An additional complementary resource is the GTeX project that aims at investigating the tissue specific mechanisms of gene regulation. This project seeks to identify expression quantitative trait loci (eQTLs), which are genomic regions containing variants that influence the expression level of one or more genes [96]. Furthermore, variants could be annotated with respect to non-coding RNAs (miRNAs and lincRNAs) location. Non-coding RNAs represent a class of transcripts that are not translated into proteins but have a regulatory role, thus potentially affecting gene function.

Immune-mediated diseases and autoimmune diseases

Among the several diseases spontaneously affecting dogs and humans, immune-mediated disorders represent a class of pathologies still difficult to completely decipher. Immune-mediated diseases derive from functional perturbations of the immune system that can either mount a disproportionate response (inflammatory diseases) or lose the ability to recognize "self" cells and tissues thus reacting against them (autoimmune diseases). Nevertheless, these diseases in some cases do not form defined clusters, but rather represent a continuum with variable overlapping genetic determinants and clinical manifestations [97]. Autoimmune disorders can be further classified based on where the target (antigen) of the aberrant immune response is expressed. Systemic autoimmune diseases typically result from the formation of circulating immune complexes constituted by nuclear proteins and specific autoantibodies. Organ-specific autoimmune diseases, in contrast, derive from autoantibodies or autoreactive T-lymphocytes responding to antigens expressed only in particular tissues [98].

During an immune response against a pathogen, antigen-presenting cells (APCs) present antigens in a major histocompatibility complex (MHC)-restricted manner to naïve helper T-lymphocytes, which subsequently get activated and result in their clonal expansion and differentiation into effector T-cells. Effector helper T-cells produce specific cytokines and express surface molecules, which in turn prompt macrophages to erase the antigens and B-cells to produce antibodies. After pathogenic antigen presentation by APCs, naïve cytotoxic T-lymphocytes become effectors and kill the cells expressing the target antigens. In autoimmune diseases, the immune response targets self-antigens. An intricate multifactorial framework contributes to the breakdown of self-tolerance and to the occurrence of autoreactive T- and B-lymphocytes. Central tolerance, which is the mechanism responsible for the ability to discriminate between self and non-self, is induced in the thymus and bone marrow by the high affinity interaction between immature lymphocytes and self-antigens. Conversely, peripheral tolerance occurs in lymph nodes and causes unresponsiveness to self-antigens expressed in the

periphery and attenuated response to environmental molecules. Several genes, including *AIRE*, *CTLA-4* and *PD-1*, as well as different cellular mechanisms regulate the acquisition of self-tolerance. Autoimmunity might derive from a failure of any mechanism controlling central or peripheral tolerance [98].

In humans, autoimmune diseases represent a major medical burden and globally affect approximately 1 in 25 individuals [99]. Strikingly, a clear sex bias exists for these diseases, with females having higher susceptibility than males [100]. Many hypotheses have been proposed to explain sexual dimorphism in autoimmunity, including fetal microchimerism, sex hormones and their role in self-tolerance, as well as X chromosome inactivation and its gene dose effect [101-103]. However, the mechanism for this female bias is still poorly understood. Autoimmune diseases cluster in families and show high concordance in twin studies, overall suggesting the presence of a major underlying genetic component. Most autoimmune diseases are polygenic disorders that develop in susceptible individuals that inherit multiple risk genetic variants; however, they are characterized by an intricate cross-talking between the immune system and the environment. A simple hypothesis is that susceptibility genetic polymorphisms result in faulty regulation of the immune response mechanisms and environmental factors initiate or augment the activation of lymphocytes reacting against self-antigens. Many environmental factors have been proposed as autoimmunity triggers, including microbial infection, occupational exposure to certain harmful molecules, vitamin D and tobacco smoke [104-106]. Several risk loci are shared between autoimmune diseases, suggesting pathogenic mechanisms affecting general immune regulation and self-tolerance. The common genetic susceptibility is also consistent with the co-occurrence of different autoimmune diseases in the same individuals and families. However, disease-specific genetic associations suggest the presence of unique biological mechanisms underlying the full spectrum of autoimmune diseases [5, 99, 107-109].

The variation at the MHC significantly contributes to disease-specific genetic susceptibility [98]. The human leukocyte antigen (HLA) complex is a particularly gene-dense region characterized by high genetic variation and long-range LD. HLA genes were the first to show association with autoimmunity and for most autoimmune diseases still represent the loci explaining the greatest fraction of the disease genetic variance [5, 110]. Several autoimmune diseases are typically associated with HLA class II alleles. HLA class II proteins are expressed in specialized APCs and participate in the selection and activation of helper T-cells, which in turn regulate the immune response against protein antigens, thus suggesting a primary role in self-tolerance mechanisms and regulation. In contrast, many seronegative inflammatory diseases are usually associated with HLA class I alleles, which are often disease-specific. HLA class I proteins are expressed in almost all

nucleated cell-types and participate in the activation of cytotoxic T-cells [98].

According to our current knowledge about most complex diseases, a large fraction of the genetic variance underlying autoimmune disease susceptibility needs to be explained. Despite the establishment of large samples cohorts and great improvements in sequencing technologies, genome annotations and analytical tools, it is currently unclear at which extent common and rare variation, including structural changes, might contribute in solving the "missing heritability" dilemma. Dogs might be of great help in this situation [69, 111-113]. Disease genetic studies in dogs could identify loci potentially causative or being involved in the same pathogenic pathways as the human counterpart of the disease, thus providing novel insights and improved understanding of the genetics of human autoimmunity. Nevertheless, this knowledge might also promote dogs' health through the development of novel diagnostic tools and targeted breeding strategies.

Hypothyroidism, a disease shared by humans and their best friends

Human hypothyroidism

Hypothyroidism is one of the most common endocrine diseases affecting humans. In this disorder the thyroid gland fails to produce sufficient amounts of thyroid hormones (Thyroxine or T4 and Triiodothyronine or T3) [114]. Low concentration of thyroid hormones causes an increase of thyroid-stimulating hormone (TSH) levels through a negative feedback mechanism. TSH induces the thyroid follicular cells to synthesize thyroglobulin (Tg), which reacts with iodine in the glandular colloid space to produce T4 and T3. T4 and T3 are released into the blood after proteolysis. Apart from the small and active fraction of hormone not bound to transport proteins (free T4), a large amount of T4 is then converted into the active molecule T3 within the target cells. Thyroid hormones primarily control the regulation of metabolism. Hypothyroidism symptoms are generally non-specific, including tiredness, weight gain and poor ability to tolerate cold, reflecting the key function of thyroid hormones in tweaking the metabolism of the body. Moreover, symptoms can vary from being completely absent in asymptomatic individuals to be extremely severe and cause a generalized multisystemic failure in the worst cases. This can make hypothyroidism to be an elusive disease, difficult to diagnose at first instance [115, 116]. However, after diagnosis, hypothyroidism treatment is specifically addressed (replacement treatment with synthetic thyroxine) and able to ensure a good life-quality. Different levels of classifications exist for this disease, depending on its specific time of onset, the anatomical location of the endocrine pathology and its severity. Hypothyroidism could be congenital, if it is already present at

birth, or acquired if it develops after birth under the influence of both genetic and/or environmental factors. Endocrinologists also refer to primary hypothyroidism if the thyroid gland itself is the affected organ and to central hypothyroidism if the hypothalamic-pituitary-thyroid axis is defective and leads to a suboptimal tuning of the hormones regulating the thyroid function. Lastly, depending on how severe the disease manifestations are, this disease could be classified as overt or subclinical hypothyroidism [116].

The synthesis of functional thyroid hormone is dependent on iodine. In geographic areas with insufficient natural and supplementary iodine, its scarcity is the major cause for the development of congenital, but also acquired, hypothyroidism [117]. However, in the developed countries where iodine supplementation or its intrinsic presence in the environment is sufficient, congenital hypothyroidism is a rare and sporadic disorder. Cases of congenital hypothyroidism have also been associated with mutations in genes encoding proteins and transcription factors participating in the thyroid function and regulation [116, 118].

In western countries, epidemiological surveys have estimated the prevalence of hypothyroidism to be around 2-5% in the general population; autoimmune hypothyroidism, defined as Hashimoto's thyroiditis (HT), accounts for the great majority of these cases [116]. HT generally occurs more often in women than in men, who have around seven-fold lower risk. Furthermore, HT incidence increases during middle age. Categories of individuals with an increased risk of developing HT are post-partum women and individuals with a familiar history of HT or other autoimmune disorders (*e.g.* type 1 diabetes, autoimmune Addison's disease, vitiligo, coeliac disease and Sjögren's syndrome) [109, 116, 119, 120]. Together with HT, Graves' disease (GD) is included in the organ specific autoimmune disorder broadly defined autoimmune thyroid disease (AITD) [121]. Briefly, GD is characterized by an increased production of thyroid hormone due to the presence of autoantibodies against thyroid-stimulating hormone receptor (TSHR), resulting in thyroid hormone overexpression. The overproduction of thyroid hormone leads to clinical symptoms including goitre, increased metabolism rate, weight loss and exophthalmos. Therapies with anti-thyroid drugs, as well as destruction of the thyroid gland by using radioiodine or thyroidectomy, are common treatments for GD. Therefore, originally hyperthyroid individuals may also be dependent on lifelong supplementation with thyroid hormones.

In HT, autoantibodies primarily against thyroid peroxidase (TPO), but also against Tg, could be found in almost all the cases. However, a small fraction of patients could be characterized by complete absence of autoantibodies, which could represent an end-stage of the disease. Moreover, autoantibodies to thyroid antigens might be detected before the actual onset of HT and be present in clinically healthy individuals. This is probably due to individual overall biological variability, as well as to the extreme complex nature of the disease, but no consensus exists regarding the possible reasons for

such physiological exceptions [116, 122]. HT develops when the self-tolerance to specific thyroid proteins is broken. In this context, autoantibodies against the thyroid molecules are produced and thyroid is infiltrated by lymphocytes and other immune-cells, thus leading to the gland destruction [116, 123]. Although many hypotheses have been proposed, the cause for the self-tolerance loss is still unknown [119, 123]. HT diagnosis firstly relies on the evaluation of a number of clinical signs, which include a wide range of dermatological, metabolic and behavioural modifications, in addition to a basic anamnesis and familiar history assessment. The clinical examination is always supported by biochemical tests, in which characteristic serological parameters reflecting the thyroid function are measured. Laboratory analysis targets include TSH, free T4, autoantibodies against TPO and Tg, and in some cases T3 that may help in the alternative diagnosis of non-thyroidal illness [116, 124].

Disease family history and disease concordance in twin studies indicate a substantial genetic susceptibility underlying HT [125-128], with the latter providing at the same time evidences of an important role of the environment in triggering the disease development. Specific environmental risk factors for HT include iodine and selenium intake, as well as Interferon alpha treatment [123]. Genetic studies have shown that certain polymorphisms in thyroid-specific genes either confer susceptibility to AITD or uniquely predispose to one of the AITD clinical manifestations (*i.e.* HT and GD). One of the genes associated to AITD is *TG*, which encodes the protein thyroglobulin [129]. Another thyroid-specific gene, *TSHR*, has been uniquely associated to GD [130]. Conversely, other polymorphisms located in thyroid non-specific genes increase the risk of general autoimmunity, being associated with AITD and other autoimmune diseases. Among this category, candidate gene approaches were able to identify associations with HLA class II (HLA-DR3, HLA-DR4 and HLA-DR5) [131-134], *CTLA4* [135] and *PTPN22* [55, 136], which all participate in the immunological synapse. *CTLA4* (cytotoxic T-lymphocyte antigen 4) plays a role in inhibiting T-cell activation, whereas *PTPN22* (protein tyrosine phosphatase, non-receptor type 22) is involved in T-cell signal transduction. Other genes in this category are *IL2RA* (interleukin-2 receptor alpha) [137], *FOXP3* (forkhead box P3) [138] and *CD40* [139]. GWA studies have confirmed known AITD susceptibility loci, as well as identified novel major associations, including, *FCRL3* (Fc receptor-like protein 3) [140, 141] and *HLA* class I [141]. The Immunochip has given an additional boost to the search for additional AITD susceptibility loci and has led to the detection of *BACH2* (BTB Domain And CNC Homolog 2), *MMEL1* (Membrane Metalloendopeptidase Like 1), *TRIB2* (Tribbles Pseudokinase 2), *LPP* (LIM Domain Containing Preferred Translocation Partner In Lipoma), *PRICKLE1* (Prickle Planar Cell Polarity Protein 1) and *ITGAM* (Integrin Subunit Alpha M) [142]. A more recent GWA study meta-analysis reported *MAGI3* (Membrane Associated Guanyl-

ate Kinase, WW And PDZ Domain Containing 3) as being associated with AITD [143]. Nevertheless, according to the NHGRI-EBI catalog of published GWA studies, *VAV3* (Vav Guanine Nucleotide Exchange Factor 3) is the only gene statistically significantly and uniquely associated with HT [144].

Canine hypothyroidism

In dogs, hypothyroidism is a very common disorder, which in almost all the cases manifests itself as primary hypothyroidism (*i.e.* the thyroidal gland being directly affected) [145, 146]. Mutations leading to a rare congenital hypothyroidism have also been described in Toy Fox and Tenterfield Terriers [147, 148], but this disease form is extremely sporadic and uncommon. Hypothyroidism is in the great majority of the cases represented by canine lymphocytic thyroiditis, also known as autoimmune hypothyroidism [149-151]. Autoimmune hypothyroidism is described as the immune-mediated destruction of the thyroid gland after the invasion of B- and T-lymphocytes, which eventually leads to the loss of thyroid function and overt clinical signs. Thyroglobulin autoantibodies (TgAAs) are present as major determinants of autoimmunity [149-152]. Canine hypothyroidism might also be caused by thyroid idiopathic atrophy, which is characterized by a degenerative nature rather than an autoimmune event. Nevertheless, it was hypothesized that this atrophic form could represent the end stage of autoimmune hypothyroidism [146]. Hypothyroidism is a widely-spread disorder in dogs, with several breeds having an increased risk of developing it. According to different health surveys and epidemiological studies, several medium-large size purebred dog breeds have been suggested as highly predisposed to develop hypothyroidism. These include the Beagle, Boxer, Doberman Pinscher, English Setter, Giant Schnauzer, Gordon Setter, Hovawart, Old English Sheepdog and the Rhodesian Ridgeback [152-157]. The disease also notably occurs within families, overall suggesting the presence of hereditary components [158]. The genetics underlying canine hypothyroidism has not been studied extensively, and only a few heritability estimates have been reported. These include values equal to 0.2 – 0.3 in the Beagle [153, 159] and approximately 0.5 in a Finnish population of Hovawart [160]. Canine hypothyroidism represents a promising model for human HT because of shared clinical manifestations [152], biochemical modifications and disease progression characteristics [146]. It is worth mentioning that hyperthyroidism is rare in dogs, thus suggesting the presence of unique genetic determinants for hypothyroid disease in this species.

The dog leukocyte antigen (DLA) class II gene cluster was the first locus found to be associated with canine hypothyroidism [161, 162]. However, similarly to the scenario seen in humans, variability at the DLA class II genes does not account for all the susceptibility of the disease in all the breeds. This suggests a complex nature of the disease, with several genetic

factors involved. Furthermore, each DLA variant moderately contributes to the overall risk of developing the disease [146]. In contrast to human hypothyroidism, which has been genetically characterized with the discovery of additional loci associated with the disease susceptibility, very little is known regarding the genetics behind development of canine hypothyroidism. Consistently with human HT description, yet unknown environmental triggers have also been included in the canine disease description. Canine hypothyroidism represents an insidiously progressive disease, with diagnosis not always immediate. However, when correctly diagnosed, canine hypothyroidism is promptly treated with a replacement therapy based on synthetic thyroxin. In dogs, the diagnosis of hypothyroidism is based on similar phenotypic and serological parameters as those in humans. Characteristic behavioural (lethargy, depression), dermatological (hair loss, dry skin, poor fur quality) and metabolic changes (weight gain, weakness, cold intolerance) are evaluated together with the measurement of peculiar functional thyroid-specific circulating molecules, such as TSH, free T4 and TgAA [146].

Additional insights into canine hypothyroidism genetics and etiology are highly desirable, even though the disorder is easily treated with artificial thyroid hormone replacement therapy. One of the reasons is the possibility to develop a genetic test to be used in breeding practices, especially before the dogs' breeding debut. Testing dogs at high risk of developing canine hypothyroidism prior to their use in breeding would allow breeders to select those dogs with lower disease genetic susceptibility even before the manifestation of any clinical signs, eventually resulting in the elimination of the risk allele(s) from the general population. Moreover, novel loci that increase the susceptibility of canine hypothyroidism could also potentially be shared with humans and could explain a fraction of the missing heritability that characterizes the human counterpart of this disease.

Autoimmune Addison's disease

Human autoimmune Addison's disease (AAD) is a rare and potentially life-threatening endocrine disorder deriving from the autoimmune destruction of adrenal cortex cells, which is the major cause of primary adrenal insufficiency [163]. However, other causes that can lead to primary adrenal failure include tuberculosis, human immunodeficiency virus (HIV), haemorrhage and metastatic malignancies. In AAD patients, the cortical cells of the adrenal glands fail to produce their characteristic hormones derived from cholesterol: cortisol that regulates stress management and immune response, aldosterone that regulates blood pressure and the excretion of specific mineral ions in the kidneys, as well as the androgenic hormones androstenedione and dehydroepiandrosterone [164]. If AAD is not treated with cortisol replacement, it may result in the so-called Addisonian crisis, which is a condition of severe adrenal insufficiency that could be eventually lethal [165].

The prevalence of AAD in Caucasians is approximately 100 per million, with differences among distinct geographic areas. Moreover, AAD incidence has been increasing during the last decades [166-172]. Women have a higher risk of developing AAD, which generally appears at middle age in most of the cases [167, 173]. AAD symptoms are mainly non-specific, including fatigue, abdominal pain, hypotension and nausea. However, characteristic hyperpigmentation, salt craving and loss of libido in women represent symptoms described in most of the patients [174]. In AAD patients, the autoimmunity hallmark is represented by the presence of the autoantibodies against 21-hydroxylase (21-OH), which are detected in 86% of the cases [165, 167, 174]. AAD generally progresses through different stages: in the earliest stage patients are characterized by the presence of the specific autoantibody, showing neither adrenal-specific dysfunction nor clinical signs. In the following sub-clinical stage clinical manifestations are still absent but the adrenal function is progressively affected, thus causing decreased adrenocortical hormone production and subsequent increased level of circulating adrenocorticotropin [175]. In clinical AAD, the disease symptoms appear when the adrenal function is almost completely compromised. AAD diagnosis generally relies on the presence of autoantibodies against 21-OH [176].

An extremely complex phenotype characterizes AAD. This disease could manifest itself either as isolated AAD, or as a part of an autoimmune polyendocrine syndrome (APS) [177]. APS-1 is a rare disorder, with varying prevalence in different populations. APS-1 is a Mendelian disease with an autosomal-recessive mode of inheritance, resulting from mutations in the *AIRE* (Autoimmune Regulator) gene. APS-1 clinical picture mainly includes adrenal insufficiency, chronic mucocutaneous candidiasis and hypoparathyroidism. Autoantibodies are also detected in APS-1, but they are mainly against a mitochondrial cytochrome P450 enzyme that acts on cholesterol [174]. Approximately half of AAD patients simultaneously develop AITD and/or type 1 diabetes, which is described as APS-2 [178]. Besides AAD, APS-2 patients might develop other autoimmune disorders such as pernicious anaemia, vitiligo and primary hypogonadism. Isolated AAD and AAD in the APS-2 context may be considered as the same entity. They have a complex genetic nature, in which different genes and combination of susceptibility variants, in addition to environmental factors, are likely to play a role in disease development [175, 179].

AAD shows high concordance rates in twin studies and aggregation in families, thus suggesting a strong genetic contribution underlying its development [180-185]. Moreover, the different autoimmune diseases clustering in individuals affected by AAD and in their relatives suggest the presence of shared susceptibility loci for these disorders. Apart from the well-established associations with the HLA complex, the loci that have been further associated with complex AAD underlie autoimmune diseases comorbid with AAD, thus confirming the potential presence of a shared etiology. These loci in-

clude, *CTLA4*, *PTPN22*, *FCRL3*, *CIITA* (Class II Major Histocompatibility Complex Transactivator), *NLRP1* (NLR Family Pyrin Domain Containing 1), *VDR* (Vitamin D receptor), *CYP27B1* (Cytochrome P450 Family 27 Subfamily B Member 1) (102half), *STAT3* (Signal Transducer And Activator Of Transcription 3) and *GATA4* (GATA Binding Protein 4) [179, 186]. However, these associations explain only a modest proportion of the genetic contribution to AAD. The detection of these susceptibility loci has relied on a candidate gene approach in small sample cohorts, thus not allowing an independent, highly powered and completely unbiased genomic scan of the entire genome. It is therefore desirable to take up such studies using larger cohorts of AAD patients, which might be difficult for such a rare disease, unless long-term projects are financially supported.

Dogs spontaneously develop AAD, which has the same clinical manifestations as the human counterpart of the disease. Dog breeds with high AAD prevalence include the Bearded Collie, Nova Scotia Duck-tolling Retriever, Portuguese Water Dog and the Standard Poodle [187]. The high disease susceptibility in these breeds and the characteristic aggregation in families strongly support a genetic aetiology for canine AAD. Pedigree studies have estimated disease heritability equal to 0.49 in Portuguese Water Dog, 0.76 in Standard Poodle and 0.98 in Nova Scotia Duck-tolling Retriever [188-190]. *DLA* and *CTLA4* have also been associated with canine AAD, suggesting an immunogenetic aetiology shared with humans [191, 192]. However, no specific autoantibodies against adrenal gland molecules (for example against 21-OH) have been identified [193], thus suggesting a heterogeneous nature of canine AAD, with some breeds having the autoimmune form of the disease and others being characterized by developmental defect of steroidogenesis. Recently, Friedenberg and colleagues [194] failed to detect a single, major locus associated with canine AAD after performing a well-powered GWA analysis (61 cases and 72 controls) in the Standard Poodle, thus confirming the hypothesis of canine AAD being highly heterogeneous. Nevertheless, the fixation of the AAD causative loci in the Standard Poodle could alternatively explain the negative outcome of such GWA analysis. Overall, this might suggest that genetic studies of human AAD could be useful for research in dogs, but not vice versa.

Aim of the thesis

The overall goal of this thesis was to expand our knowledge regarding the genetic background of immune-mediated diseases. The research presented here aims at unraveling novel and confirming known genetic loci underlying two immunological diseases by using a disease animal model, the domestic dog, as well as humans.

In particular, specific goals were the following:

- To investigate canine hypothyroidism as a model for human Hashimoto's thyroiditis in high-risk dog breeds in order to identify new and confirm known genetic factors underlying the disease (Paper I and Paper II).

- To identify novel and confirm established genes associated with the development of autoimmune Addison's disease in humans (Paper III).

Present investigations

Papers I and II: Canine hypothyroidism as a model for human Hashimoto's thyroiditis: disease mapping in high-risk dog breeds

Background

Disease phenotype, prevalence and diagnostic considerations

Canine hypothyroidism is a complex clinical condition and represents one of the most common endocrinopathies in the domestic dog. If we exclude rare cases of congenital hypothyroidism characterized by dwarfism and/or goiter, as well as other sporadic thyroid disorders such as central hypothyroidism, primary hypothyroidism accounts for the great majority of cases of this disease in the domestic dog. The most common cause of primary hypothyroidism is lymphocytic thyroiditis, which is characterized by the progressive autoimmune destruction of the thyroid gland. Thyroid idiopathic atrophy has been recognized as an alternative cause of primary hypothyroidism, even though it might reflect the final stage of autoimmune hypothyroidism. Several dog breeds have an increased susceptibility to hypothyroidism compared to the general dog population in which the disease prevalence has been estimated to be lower than 1% [195]. For example, epidemiological studies have shown a significantly higher prevalence of hypothyroidism in the Swedish Hovawart (13%) and Giant Schnauzer (16%) cohorts, as well as in the Beagle (16%) and in the Gordon Setter (3%) [153, 195-197]. Other breeds, such as the Rhodesian Ridgeback, have also been described with significantly increased prevalence of hypothyroidism specific autoantibodies [149]. Furthermore, hypothyroidism typically aggregates in pedigrees, overall suggesting a major role of genetic determinants in disease etiology.

Similarly to dogs, the majority of the primary hypothyroidism cases in humans derive from autoimmune hypothyroidism, also known as HT [198]. HLA has been associated with autoimmune hypothyroidism in humans, and correspondingly DLA has been associated with increased susceptibility to hypothyroidism in high-risk dog breeds. Moreover, canine and human hypothyroidism show striking similarities, including disease etiology, clinical symptoms and disease progression, thus corroborating the effectiveness of the domestic dog as a spontaneous animal model [146, 151].

In general, the onset of canine hypothyroidism appears at middle age, although in high-susceptibility breeds the highest risk is at a younger age [199]. The clinical symptoms of canine hypothyroidism are usually non-specific and often involve a wide range of indefinite abnormalities, thus complicating disease diagnosis. Serological measurements of thyroid-specific molecules are routinely employed in veterinary clinics for the diagnosis of hypothyroidism. However, there are other factors that might affect the concentration of these molecules, thus complicating the diagnosis. In dogs, the diagnostic test assessing TSH has high specificity but low sensitivity [200]. High concentrations of TSH can confirm the diagnosis of hypothyroidism, but normal levels of TSH cannot exclude the disease [200, 201]. However, it has been shown that TSH concentrations are more likely to remain within reference range in non-thyroidal illnesses [202, 203]. A fraction of hypothyroid dogs might have the concentration of T4 within reference range because of the presence of autoantibodies against T4; furthermore, T4 concentrations might decrease as a result of aging, as well as certain non-thyroidal diseases and drugs. Conversely, free T4 levels are less affected by extra-thyroidal perturbing factors [146, 202, 204, 205]. The serological test for TgAA is characterized by low sensitivity but in general good specificity. This test usually confirms the diagnosis of autoimmune hypothyroidism, although these autoantibodies might be present even before clinical disease manifestations and absent during the disease end-stage [146, 149, 206].

For all these reasons, the diagnosis of hypothyroidism might be a true challenge, thus requiring, when possible, a comprehensive diagnostic panel comprising of thyroid specific serological measurements and expert evaluation of clinical signs.

Study design

Canine hypothyroidism has not been extensively dissected from a genetic point of view. Genetics clearly plays an important role in the hypothyroidism etiology as demonstrated by the extent of disease aggregation in pedigrees and breed-specific disease susceptibility. However, very few genetic studies have been carried out. For instance, potential factors hampering genetic studies in dogs might be sample collection and presence of considerable cryptic relatedness, as well as phenotyping when the disease shows extreme clinical complexity. What we currently know about the genetics of canine hypothyroidism results from a few candidate gene studies that have detected associations with DLA in a number of dog breeds [155, 161, 162, 197, 207]. However, genetic variation at this locus explains neither the entire underlying risk in specific dog breeds nor the genetic susceptibility shared among breeds. Furthermore, candidate gene studies show intrinsic limitations that might affect the analytical outcomes.

Evidently, there is a need for unbiased studies based on genome-wide scans in order to gain insights about the genetics of canine hypothyroidism.

When the sequencing of the dog genome was completed and the foundations of the current canine genomic research were being built, it was hypothesized that a GWA study of 100 cases and 100 controls would be sufficient to reliably map an allele conferring a fivefold risk to develop a complex disease [31]. However, researchers have successfully mapped osteosarcoma, systemic lupus erythematosus (SLE)-related disease, cardiomyopathy and atopic dermatitis using fewer than 200 dogs [27, 113, 208, 209].

NGS technology allows the detection of all polymorphisms present in coding and non-coding regions, thus representing a great analytical support to the current well-established, cost-effective and time-efficient GWA mapping approach. NGS indeed facilitates fine mapping and permits the eventual discovery of potential causative mutations, which is the ultimate goal of genomic screenings.

Paper I: A shared susceptibility locus for canine hypothyroidism identified through a multi-breed genome-wide association analysis

Study samples and phenotyping

Blood and serum samples were collected from three dog breeds with high susceptibility to develop canine hypothyroidism. These three breeds were the Gordon Setter (GS), Hovawart (HV) and the Rhodesian Ridgeback (RR). GS samples (n=165) were collected in Norway, HV samples (n=74) were obtained from different European countries, whereas RR samples (n=92) were all collected in the United States of America. All dogs were categorized into cases or controls according to clinical diagnoses evaluated by expert veterinarians. Further, the clinical disease phenotype had to be supported by thyroid specific serological measurements, such as increased levels of TSH (> 40 mU/L) and reduced levels of free T4 (< 7 pmol/L), whereas controls had to be older than seven years. Cases with other potentially impairing conditions, as well as controls with other immunological diseases, were excluded after a review of clinical data and/or questionnaires completed by the dog owners. Data were harmonized according to shared diagnostic criteria, since the samples presented distinct analytical spectra and were based on specific reference ranges according to national laboratory standards.

Analysis and results

We genotyped all case and control samples using ~170,000 SNP markers (Illumina 170K HD BeadChip) and, after data quality control (QC), we performed an association analysis in each dog breed separately. In every GWA analysis, we employed a mixed model approach that is able to correct for both population structure and cryptic relatedness. Considering that HV samples had different geographic origins, we wanted to assess whether this

would have resulted in the presence of any structure in the population. For HV, we therefore created a vector reflecting the most probable number of clustering sub-populations and we used this vector in the mixed model. In order to test the statistical significance of the breed-specific association analyses, we set different significance levels: empirical genome-wide significance after 1,000 permutations and empirical 95% confidence intervals of the markers distribution. The multi-breed GWA analysis detected a common peak of association on *Canis familiaris* autosome (CFA) 12 (Paper I, Figure 2a-2c). We found a statistically significant association in GS and HV, whereas a suggestive association in RR (Paper I, Figures 1d-1f and 2a-2c). The top associated SNP effect had the same direction and approximately the same magnitude in all the breeds: the odds ratios (ORs) were ranging from 3.4 in RR to 5 in HV (Paper I, Table 1). Population stratification was not present in any of the breeds as demonstrated by the λ values of the associations (GS=1.019; HV=0.995; RR=0.996) and also by the multidimensional scaling plots where the case and control samples clustered homogenously (Paper I, Figure 1). In each breed, we used LD to locate the regions of association based on squared coefficients of correlation values (r^2) ≥ 0.7 between each breed specific GWA top SNP and SNPs in CFA12. This resulted in the definition of a ~2.5 Mb region of association shared by the three breeds and corresponding to the associated region in HV (Paper I, Figure 3). We then decided to perform a fixed effect meta-analysis assuming the presence of a shared genetic effect being the same in the different breeds. The meta-analysis identified a top associated SNP ($p=2.1x10^{-11}$) located within the shared associated locus (Paper I, Figure 2d). After phasing the genotypes of the overlapping associated region, risk haplotypes in cases (two) and non-risk haplotypes in controls (nine) were defined in each breed separately according to the meta-analysis top SNP genotype. By identifying the recombination breakpoints of the risk haplotypes, we defined the minimal risk haplotype, and the corresponding non-risk haplotype, shared across breeds. This shared risk haplotype spans ~167 Kb and harbors three genes (Paper I, Figure 5). The minimal shared haplotype is tagged by two SNPs that were tested for association with the disease phenotype as both haplotypes and genotypes. We detected a significant enrichment of the risk haplotype, as well as the two SNPs risk genotypes, in cases versus controls, both in each breed separately and across breeds (Paper I, Table 2).

Discussion

In this study, we presented the first GWA study of canine hypothyroidism. In the past, mapping efforts focused on associating this disease to genetic variation detected in genes linked to the human disease counterpart (*i.e.* candidate gene studies). With this unbiased study we sought to identify novel genetic risk factors increasing disease susceptibility in three high-risk breeds. Moreover, we tested the possibility of a genetic locus being shared

between different dog breeds. For the first time a risk haplotype contributing to a complex disease was detected across different dog breeds and subsequently confirmed as shared via meta-analysis. Our analysis was grounded in the two-stage mapping strategy suggested by Karlsson and colleagues [210], in which a causative locus could be firstly mapped in one breed and subsequently fine mapped in a second breed sharing the same phenotype and IBD causative variant(s). However, we complemented this approach with meta-analysis, which is an alternative mapping methodology widely employed in human genetics. The integrated multi-breed GWA and meta-analysis strategy allowed the discovery of a shared ~167 Kb risk haplotype on CFA12; this haplotype was enriched across the affected dogs and emerged to be independent of the nearby DLA locus. This suggested that either this risk haplotype might represent an additional susceptibility locus for hypothyroidism, or that previous candidate gene study associations with DLA in high-risk breeds might have resulted from high LD with this risk haplotype in the used breeds.

The identified haplotype harbors three genes (*LHFPL5*, *SRPK1* and *SLC26A8*) that have not previously been associated with hypothyroidism. These findings could lead the way to the description of entirely novel pathways and mechanisms playing a role in hypothyroidism etiology. *LHFPL5* is a gene encoding for lipoma HMGIC fusion partner-like 5. Mutations in this gene cause deafness in humans [211, 212], and in mice [213]. *SRPK1* encodes a serine/arginine protein kinase specific for the serine/arginine-rich domain family of splicing factors, and it has been shown as being an important factor in tumorigenesis [214], viral infection [215, 216] and apoptosis [217]. *SLC26A8* (solute carrier family 26 member 8) is a member of the *SLC26* gene family of anion transporters. Variants in one member of this gene family, specifically in *SLC26A4*, have been shown to cause Pendred syndrome, a genetic disorder characterized by goiter and sometimes also hypothyroidism [218]. Nevertheless, it is possible that putative variants within the haplotype regulate the function of genes located outside this locus, since non-coding functional elements are found within the identified region.

The divergence between the size of the breed-specific associated loci, as well as the shared risk haplotype, and the theoretical length of the haplotypes within and across dog breeds, suggested the presence of potential selection targets in the identified risk haplotype. However, there were no domestication selection sweep signals in the risk haplotype when comparing the wolves and the dogs included in the study conducted by Axelsson and colleagues [37]. On top of that, the wolf population employed by Axelsson and colleagues [37] lacked the identified hypothyroidism risk alleles. Conversely, these susceptibility alleles have been identified in low-risk dog breeds. This overall suggested the origin of the identified hypothyroidism risk factor after domestication and prior to breed creation.

In conclusion, we demonstrated the remarkable potential of the integrated multi-breed GWA and meta-analysis methodology for the detection of genetic loci underlying complex diseases in dogs. These findings might eventually be implemented and translated in veterinary and human clinical practices in order to promote not only dogs' health, but also our wellbeing.

Paper II: The type I Interferon gene cluster is associated with canine hypothyroidism

Study samples and phenotyping

Blood and serum samples were collected from 115 Giant Schnauzer dogs after acquiring owners' written consent. Sampling was organized and carried out in collaboration with licensed veterinarians throughout Sweden. A comprehensive diagnostic panel of thyroid specific serological measurements was obtained for all study samples. More specifically, we determined serum concentrations of TSH, free T4 and TgAA. All dogs were categorized into cases or controls according to predetermined diagnostic criteria and reference ranges (**Table 1**) in conformity with earlier studies conducted by our group [162, 196]. Furthermore, cases and controls with additional immune-related conditions were excluded based on a follow-up evaluation of clinical data and/or questionnaires answered by dog owners.

Table 1. ***Diagnostic criteria used to classify case and control dogs (TgAA= thyroglobulin autoantibody, TSH=thyroid-stimulating hormone, T4=thyroxine, POS=positive, NEG=negative)***

Phenotype	Diagnostic criteria
Case	TgAA POS and/or TSH ≥ 40 mU/L
Control	TgAA NEG, TSH ≤ 25 mU/L, free T4 ≥ 5pmol/L, age ≥ 7 years

Analysis and Results

Following genotyping with the Illumina 170K HD BeadChip array, we performed a GWA analysis in 71 cases and 36 controls resulting from the data QC and filtering steps. We used a mixed model to adjust for potential confounders and two statistical significance levels (empirical genome-wide significance after 1,000 permutations and 95% empirical confidence intervals of the SNPs distribution) to evaluate the reliability of the association. We identified a locus suggestive for association on CFA11 conferring protection against canine hypothyroidism (OR=0.15) (Paper II, Figure 2). The candidate locus was defined as spanning ~8.9 Mb based on pairwise LD estimates ($r^2 \geq 0.8$) of the top SNP to the rest of the markers on CFA11 (Paper II, Figure 3). We then performed Illumina pair-end short reads whole-genome

resequencing (WGS) for individuals (n=23) representing the key haplotypes as detected by the GWA analysis. One control dog homozygous for the protective allele and two cases homozygous for the opposite non-protective allele were sequenced at high coverage (HC), resulting in 46X genome average coverage. Additionally, 10 controls heterozygous for the protective allele and 10 cases homozygous for the opposite non-protective allele were sequenced at low coverage (LC), resulting in 6.8X genome average coverage (Paper II, Table S1). After alignment and variant calling, we selected single nucleotide variants segregating in the HC samples and located in the region of association. Out of these ~13,000 selected variants, 740 were prioritized according to their functional potential (effect on the translated amino acid sequence, overlap with conserved elements and specific predicted transcripts and promoters) (Paper II, Table S2). The prioritized breed-specific variants were re-genotyped in the GWA sample set (n=96) using Sequenom technology in order to fine map the region of association. We also imputed the genotyped variants in the missing samples (n=11). We then performed a new association analysis using a comprehensive dataset composed of all samples (n=107) and all genotyped and imputed variants. We identified a statistically significant association to the same CFA11 protective locus, which was narrowed down to ~4.18 Mb (Paper II, Figures 5 and 6a). By screening the fine mapped region of association for the presence of structural variants in the sequenced individuals, we detected a putative copy number variant (CNV) consistently segregating with the identified protective locus (Paper II, Table S6). The putative CNV was detected by the CNVnator software in the HC samples (Paper II, Table 4) and was confirmed in the LC case and control groups by comparing their average normalized coverage in 1 Kb windows (Paper II, Figure 6c-d). The predicted putative CNV overlaps the upstream region of the Interferon alpha7 (*IFNA7*) gene (Paper II, Figure 6e).

Discussion

In this study, we detected a wide region of association on CFA11 conferring protection against canine hypothyroidism. The long associated locus could either derive from selection of another desirable variant that has hitchhiked the putative protective mutation, or from a recent mutational event in the Giant Schnauzer breed. The non-protective allele at the GWA analysis top SNP is fixed in the wolf population employed by Axelsson and colleagues [37], thus representing the potential ancestral allele. In connection to the results of Paper I meta-analysis, we also evaluated the CFA12 risk locus tagging SNP genotypes in the Giant Schnauzer cohort. In Giant Schnauzer, this tagging variant segregates and the risk allele shows higher frequency in controls than cases, suggesting either absence of the causative risk variant in this breed or the occurrence of recombination events between this risk tagging variant and the actual causative allele in the Giant Schnauzer dogs. Moreover, we hypothesize that the protective locus on CFA11 identified in

this study does not hide the effect of the risk variant identified in Paper I, suggesting the absence of this susceptibility locus in the Giant Schnauzer. Overall, these findings corroborated the hypothesis of a complex etiology underlying canine hypothyroidism.

In general, the high LD levels present within individual dog breeds simplify the initial mapping of a causative genetic region due to the limited analytical requirements in terms of number of individuals and markers; however, this complicates the eventual identification of the causative mutation(s). Our study demonstrated the benefit gained by the detection and the genotyping of additional segregating breed specific variants when the objective is to fine map long regions of association. This is valid when an additional dog breed sharing the same IBD candidate locus is not available, and particularly beneficial when disease mapping is performed in breeds not included in the design of the Illumina 170K HD BeadChip array, which represents the current gold standard for canine GWA studies.

The predicted putative CNV segregating with the protective locus in the resequenced samples is located in the upstream region of the *IFNA7* gene, thus potentially altering its expression. The *IFNA* gene family belongs to the type I *IFN* genes, which encode proteins with functions involved in protecting the body from viral infections and in regulating the activity of effector immune cells. Increased serum type I IFNs activity has been detected in patients with AITD [219]. Moreover, type I IFNs, particularly IFNA, have also recently emerged as key molecules in the etiology of a systemic autoimmune disease called systemic lupus erythematosus (SLE) [220-223]. A number of studies have shown a high incidence of hypothyroidism after IFNA-treatment in patients with hepatitis C virus infection [224]. Similarly, pre-existing AITD has been shown to exacerbate in response to IFNA treatment [225]. Remarkably, type I *IFNs* have been extensively associated to human autoimmune hypothyroidism and general autoimmunity.

However, due to the high genomic complexity of the associated region identified in our study, further sequencing and annotation efforts are required to validate the putative CNV and its function in the context of type I IFN activity and autoimmunity. Although significant efforts have been made in order to improve the dog reference genome (Canfam3.1) assembly and annotation [226], our study pointed out several issues characterizing Canfam3.1 in our region of interest. Some genomic regions, including our associated locus, appear to be still problematic and difficult to extricate; we thus highlight these regions as potential targets for future work to be carried out by the canine genomics community.

In conclusion, with this study we expanded the knowledge about the genetic basis of canine hypothyroidism and we posed a striking link with the human counterpart of the disease, thus confirming the effectiveness of the domestic dog as disease animal model.

Paper III: Disentangling the genetic complexity of human autoimmune Addison's disease: identification of *BACH2* as a novel susceptibility gene through targeted resequencing

Background and study design

With the advent of NGS, the study of human complex diseases has entered a stage in which it has been possible to screen the genome with an unprecedented high resolution. The Human Genome Project and the early large-scale resequencing efforts have led to the establishment of GWA studies, which permit the assessment of how common variation is associated with disease phenotypes. However, this approach is theoretically only able to guide us towards the detection of associated genetic loci rather than causative mutations. The recognition that common variants have relatively small effects on disease genetic susceptibility and are unable to entirely explain the disease genetic variance has accelerated the transition towards applications based on direct NGS practices. WGS allows geneticists to efficiently detect the whole spectrum of genetic variation at the single base-pair resolution.

When NGS applications were starting to be employed in disease genetics studies, WGS was still economically inaccessible for a large fraction of the laboratories across the world. At this stage, whole-exome sequencing (WES) represented a cost-effective alternative to WGS. In WES, all coding exons are targeted, thus allowing the detection of functional variants that have a big impact on the translated protein amino acid sequence. WES produces approximately one hundred times less data than WGS does at the same coverage, thus reducing the burden for data storage, management and analysis. Moreover, it could be strategic to sequence more samples with WES, thus gaining statistical power, rather than to sequence the complete genome of few samples. However, the big limitation of WES is intrinsic to its design. Apart from a few very recent developments of extended arrays, a typical WES design neglects all the classes of regulatory elements and non-coding genes. However, genetic variation at these loci has been strongly associated with the aetiology of complex diseases.

Even before the first reported WES application [227, 228], our group developed a sequence capture array to be used in genetic studies of immune-mediated diseases. The array was designed targeting 1853 genes involved in immune function, inflammation and autoimmunity. The list of genes included in the array was compiled based on candidate loci identified in GWA studies of immune-mediated diseases, as well as based on their known functional role in immune responses. Additional genes were selected based on a systematic evaluation of molecular pathways involved in immune function. The array also included genes causing monogenic adrenal disease. The array

targeted a few hundred genes preliminarily identified by our group in early canine disease genetics studies, thus emphasizing the possibility of promptly translating the knowledge that we gain in studies using the domestic dog into human research. The genomic coordinates of all alternative transcripts were retrieved from NCBI36/hg18 human genome assembly. For each transcript, the target encompassed the following regions of interest: 5' and 3' UTRs, coding exons, potential promoter regions identified as 2 Kb upstream and downstream the transcription start site, as well as splice sites identified as 20 bp of intronic sequences flanking the exons. Moreover, the target included a selection of potentially regulatory elements located within 100 Kb upstream and downstream, as well as overlapping the genes. These elements were defined according to 29 mammals' conservation score, using SiPhy LOD-score higher than 7 as a proxy [229].

The targeted resequencing array we developed was somewhat pioneering because its design integrates the advantages offered by both WES and WGS. Moreover, differently from the genotyping of associated variants (*i.e.* Immunochip), it allows the detection of variants located in potentially novel disease candidate genes and theoretically covering the whole spectrum of allele frequencies.

In any genetic study aiming at mapping a certain complex disease, the likelihood of successfully pinpointing a disease susceptibility locus considerably increases with the degree of phenotypic homogeneity among the patients. Although it might result in reduced sample sizes, correct phenotype definitions can enhance the power of detecting susceptibility loci due to a decreased dilution of the associated genetic effects. Autoimmune Addison's disease (AAD) diagnosis can be confirmed with a serological marker that represents the autoimmune hallmark in AAD patients, namely 21-OH autoantibodies. 21-OH autoantibodies are present in the great majority of AAD patients, and can be detected with a specificity of approximately 95%, thus making AAD patients an extremely homogenous disease group. The Swedish Addison Registry (SAR) is one of the world's largest AAD biobanks, currently comprising of more than 700 samples [166]. SAR includes a wealth of detailed clinical information and comprehensive serological data that greatly facilitate the disease phenotyping. All the AAD patients included in our study were obtained from SAR and were conforming to the diagnostic criteria for primary adrenal insufficiency. We strictly refined AAD phenotype definition of the study case group by excluding patients where other causes of adrenal failure were suspected, inclusive of diagnosed and undiagnosed APS-1, as well as patients with fragmentary clinical data. Moreover, we excluded patients lacking 21-OH autoantibodies, as well as patients lacking consecutive positive results from two independent tests for 21-OH.

Besides the degree of family clustering, a very recent twin concordance study reported a significantly high heritability value for AAD (> 0.80), further suggesting a strong genetic component contributing to the development

of this disease. However, the genetic factors underlying AAD are largely unknown. The rarity of AAD has limited the search for susceptibility genetic loci. The few loci that have been linked to AAD derive from candidate gene studies and explain only a small fraction of the disease genetic variance.

Hence, together with the targeted resequencing array we developed, SAR represents an invaluable and unique resource for undertaking large-scale and independent AAD genomic screenings.

Analysis and Results

In this study, we performed targeted NGS using Illumina short paired-end reads in extensively characterized AAD patient (700) and control samples (1501), all collected in Sweden. AAD patients were retrieved from SAR and their disease phenotypes accurately refined, whereas control samples were obtained from direct sampling of healthy individuals (n=653) or from blood donors (n=848) not suffering from AAD. In order to attain a final SNPs dataset of high quality, we applied stringent individual and variant QC methods to the data that resulted from GATK (Genome Analysis Toolkit) best practices and initial filters based on overall genotype quality. With the individual QC step we removed samples with unsatisfactory mean target coverage, as well as samples predicted to be either non-European or highly related (Paper III, Figures S1 and S2). Additionally, we discarded samples with discordance between reported and inferred sex, as well as samples based on several sequencing quality parameters that deviated five standard deviations from the mean of the whole cohort (Paper III, Figures S3-S5). We then removed samples according to their fraction of missing data. Conversely, with the variant QC we removed variants based on their proportion of missing data across the samples, as well as between case and control samples (Paper III, Figure S6). The final dataset included 479 cases and 1394 controls, as well as ~100,000 common SNPs (MAF > 1%) and ~300,000 rare SNPs (MAF ≤ 1%). The mean variant call rate (96%) and the mean individual call rate (95%) indicated a dataset of high quality. We performed an association analysis of common SNPs by using a logistic regression corrected for the multi-dimensional scaling coordinates that reflected population substructure separating samples collected in northern versus central/southern Sweden (Paper III, Figure S7). The residual inflation was subsequently corrected using genomic control that accounts for cryptic relatedness (Paper III, Figure S12). After Bonferroni correction, we identified a statistically significant association to the *BACH2* gene ($p=1.66x10^{-15}$; OR=2.01; MAF 0.46/0.29 in cases/controls), which represents a novel AAD risk locus (Paper III, Figure 1). Moreover, we found two signals of association in the HLA region (Paper III, Figure 1 and Table 1). Although *BACH2* has been previously associated with organ-specific autoimmune diseases co-morbid with AAD [142, 143, 230-233], in this study we showed that this gene is independently associated with

isolated AAD (Paper III, Figure S14). The associated ~215 Kb haplotype block in the *BACH2* locus includes variants in high LD with the leading top associated SNP and covers the 5' non-coding exons and introns of this gene (Paper III, Figure 2). Further, we functionally annotated the *BACH2* associated variants based on epigenetic data in order to facilitate the dissection of single variants' effects, which otherwise would be difficult because of the presence of LD (Paper III, Table S6). Additionally, we confirmed the previously established associations with HLA [167, 234-236] by employing *in silico* HLA typing; with this approach, HLA alleles are inferred from *de novo* assembly of sequencing reads to predefined haplotypic references (Paper III, Table 2). Finally, we performed rare variant association analysis by aggregating variants in spaces corresponding to Ensembl GRCh37 human reference genome genes boundaries. Using a variant component test, no genes were statistically significantly associated after evaluating either the whole set of rare variants or only high and moderate-impact rare variants (Paper III, Figure S16).

Discussion

In this study, we presented the first large-scale study of the genetic background of AAD. We sought to identify novel genes associated to the development of AAD by branching off from standard analytical procedures that have characterized genetic research in AAD. We expanded the investigation scope, both in terms of genomic screening targets and sampling resources. To reach this goal, we developed a sequencing array targeting a comprehensive selection of potential causative genes involved in immune function, as well as their surrounding non-coding regions. Additionally, we exploited the exceptional value of SAR, which is one of the largest collections of AAD samples accessible for investigation. On top of that, the exhaustive phenotypic characterization of these samples represented the underpinning of this study.

We identified *BACH2* as a novel risk locus in AAD. We were also able to show that the strong association signal in *BACH2* did not depend on either autoimmune comorbidity or presence of autoantibodies against TPO [143]. *BACH2* is a transcription repressor that regulates B-cells immune function, mainly the immunoglobulin heavy chain transcription, as well as the class-switching event. Although initially linked to B-cell function, *BACH2* has also been described to be a master regulator of T-cell function by modulating the balance between cytotoxic T-cells versus helper and regulatory T-cells [237-239]. Moreover, *BACH2* has been associated with other autoimmune diseases, including type 1 diabetes and AITD. For these reasons, *BACH2* emerges as a strong candidate gene in organ-specific autoimmunity. Although our findings do not directly imply any causal relationship with AAD, the *BACH2* associated 5' regulatory region comprising of functionally anno-

tated candidate variants provides a guide for further investigations and functional studies.

In this study, we further confirmed a well-established contribution of HLA to AAD susceptibility. Certain haplotypes known to increase AAD risk in other populations were found to be associated in our AAD Swedish cohort as well. Even though we selected only AAD patients showing 21-OH autoantibodies, we detected a large variability at the HLA locus, thus suggesting that homogeneity at this locus is not necessary to induce immune reactivity against 21-OH in the context of AAD.

Apart from HLA, we did not detect statistically significant associations to other known AAD susceptibility loci. This could derive from the inclusion of a too strict significance threshold in our study to avoid false-positive associations, the previous potential detection of population-specific alleles that might fail to be replicated in our Swedish population, as well as the potential spurious association resulted from candidate gene studies that cannot incorporate population substructure information from genetic data. Similarly, we did not reveal any additional candidate loci from rare variant association analysis. It is likely that the current sample size of this study limited the power of detecting rare variant contributions to AAD etiology.

In conclusion, the knowledge gained in this study adheres to the notion of regulatory variants potentially being major contributors to complex diseases, and may eventually represent a step forward towards the promotion and the development of preventive disease treatment for AAD.

Concluding remarks and future perspectives

In recent years, biomedical research has made giant strides in exploring and understanding the genetic underpinnings of complex diseases. Although a wealth of variants has been associated with a wide range of diseases, one of the most difficult tasks now is to identify disease causative mutations and gain insight into their mechanistic role in pathology development. Moreover, an equally considerable challenge is to identify the genetic fraction contributing to disease development that is still invisible to our eyes. The work presented in this thesis made an effort towards this direction. We attempted to unravel novel genetic factors underlying complex diseases with a clear immunological etiology both in dogs and humans.

In Papers I and II we described novel genetic loci associated to canine hypothyroidism.

In Paper I we identified a shared genetic risk locus enriched in three dog breeds with high susceptibility to canine hypothyroidism. However, no causative mutation has been pinpointed. For this purpose, NGS of dogs with and without the risk haplotype would be desirable. NGS would allow identifying variants that will be subsequently functionally annotated to restrict the list of candidate mutations to be tested for biological functions. Notably, the choice of these experiments largely relies on the candidate mutation functional category. The hunt to pinpoint the putative causative mutation could be greatly aided by gene expression analysis of the three genes harbored by the identified risk haplotype. The optimal situation would be retrieving appropriate cell types of case and control dogs representing all the alternative risk haplotype genotypes and performing a comprehensive parallel expression analysis. However, availability of the complete set of samples might be limited. In addition to that, comparing case and control dogs might be difficult, because expression in case dogs could be influenced by medical treatment, thus not being trustworthy. A possible alternative could be employing control dogs with different haplotypes and scrutinize any difference in gene expression. Moreover, allele-specific gene expression analysis in control dogs heterozygous for the risk haplotype could provide information about the differential expression rate of the two alleles. This might suggest the presence of a regulatory mutation influencing gene expression, thus indicating the potential role of this locus in the development of hypothyroidism. Unfortunately, the collection of the most relevant tissue (*i.e.* thyroid) is possible only during a

surgery or post mortem. Thyroid tissue could also be obtained by biopsy, but this clinical practice might have implications on the dog welfare. Since canine hypothyroidism is in the great majority of the cases of autoimmune nature, another suitable cell type could be any collection of immune cells, for instance lymphocytes. If the putative susceptibility mutation is within the detected haplotype and affects one of the genes present within its boundaries, expression analysis could address downstream analysis towards the most plausible candidate. However, the putative susceptibility mutation could be a regulatory variant influencing the expression of a gene outside the detected risk haplotype. In this case, whole-transcriptome RNA sequencing experiments could be employed.

The detection of the risk locus on CFA12 described in Paper I could open completely new endeavors for improved canine diagnostics to be implemented along with appropriate breeding strategies, with the ultimate goal of removing the hypothyroidism risk allele from the population. Nevertheless, this finding can inform about the genetics underlying the development of HT and uncover new research paths.

In Paper II we identified a genetic locus that confers protection against canine hypothyroidism in a dog breed with high disease susceptibility. We detected a putative structural event upstream of *IFNA7* that represents an attractive variant to follow up in order to confirm its involvement in decreased disease susceptibility. This gene belongs to the type I *IFNs* gene family. Type I *IFNs* play a fundamental role in protecting the body from viral infections, thus representing an advantageous target for selection. For this reason, it would be of interest to compare genomic data from wolves and dogs in order to test the presence of a domestication selection signal overlapping this genomic locus. Moreover, it would be desirable to evaluate this structural event in resequencing data from the whole Giant Schnauzer cohort we used in this study and in other dog breeds, especially in those with high hypothyroidism susceptibility. To build up on this, it would be of primary importance to identify the precise variant breakpoints, which would facilitate functional studies to prove its biological significance in the context of canine hypothyroidism. An attractive follow up functional study would then be to reproduce this structural variant in suitable cell types by using for example the CRISPR/Cas9 technology and assess if the disruption of a potential regulatory element leads to a detectable phenotypic change. What we don't know is whether the reference genome or the dogs included in this study might bear additional rearrangements in the locus that we identified. To solve the mystery hidden within this challenging region of high complexity, it would be thus advisable to perform resequencing of bacterial artificial chromosome (BAC) clones or WGS using long reads technology. Further, the results of this study suggest measuring concentrations of IFNA in the serum of dogs with and without the identified putative variant by using a specific ELISA canine assay. An alternative approach is to undertake gene expression study

assessing *IFNA7* expression level in comparison to a housekeeping gene, taking also into consideration the high levels of sequence identity characterizing the *IFNA* gene family. RNA sequencing could also be employed to evaluate gene expression, but this experiment relies on optimal reference assembly and annotation. The findings shown in this study strikingly link canine hypothyroidism to the human counterpart of the disease, and thus corroborate the important role of type I IFN genes as candidates in autoimmunity. Moreover, the knowledge gained in this study might contribute to the development of breeding strategies, via the adoption of a marker-assisted selection eventually increasing the frequency of the protective allele in the population.

Both Papers I and II also advocate an effort from veterinarian clinicians to define standard inclusion or exclusion criteria regarding the canine hypothyroidism phenotype, which should be shared across all the veterinarian hospitals across the world. These criteria should ideally include a full and expert evaluation of disease clinical signs and always supported by the comprehensive spectrum of thyroid specific serological measurements. It would also be highly desirable to aid the above-mentioned diagnostic criteria with thyroid ultrasonography and histological examinations, although their routine implementation might be difficult. These techniques might be useful in confirming the hypothyroidism diagnosis and revealing lymphocytic infiltration of the thyroid, especially when the TgAA could not be detected during the disease end-stage.

With the studies presented in papers I and II we provide geneticists and clinicians with novel findings to be possibly translated into human research, with the hope that they will be useful for enlarging our understanding of disease etiology and for improving disease management.

In Paper III we identified a novel locus associated with AAD. We also provided a list of potential regulatory variants located in the *BACH2* gene that warrant future functional assessment. The evaluation of the identified regulatory variants and their function employing wet-lab experiments should be prioritized in order to build on our findings. For instance, possible experiments could involve electrophoretic mobility assay (EMSA) and luciferase assay in appropriate cell types. However, gene expression is specifically regulated in different cell types, at determined times and under specific environmental stimuli, thus making functional validation a true challenge. Moreover, it is of primary importance to replicate the association of these variants in other populations, thus confirming the role of *BACH2* in AAD development. Moreover, it would be of great interest to create an international consortium with the purpose of merging the different AAD cohorts and screening for additional disease susceptibility loci. This joint venture would result in a significant increase of sample size and subsequently in a higher detection power. Furthermore, WGS would be a desirable follow up study, espe-

cially considering its steadily reducing costs. WGS would provide genetic insights regarding the plethora of genes and non-coding regions not included in our sequencing array, as well as enable a basic assessment of the potential role of structural variation in AAD susceptibility. In the long run, however, it is likely that this technology will be adopted as standard method for disease mapping, making also all alternative genotyping methodologies redundant. The AAD cohort used in this study constitutes an excellent material for dissecting the complex genetics of autoimmune diseases. Among AAD patients, we can distinguish and subgroup different classes of secondary autoimmune disorders and serological categories. This suggests performing a comprehensive crosscheck in order to discriminate between different variants' genetic effects. The findings shown in this study, appropriately supported by future follow up experiment can boost research towards better disease treatment and represent a step forward towards the eventual realization of personalized medicine.

Nevertheless, it would be of extreme value to explore the function of *BACH2* in dog breeds with high susceptibility to canine AAD. We propose *BACH2* as an attractive target for candidate gene studies in canine AAD, which might partially help overcoming the difficulties in unravelling the genetic background of this disease in dogs.

In conclusion, the knowledge that we gain from genetic studies in the domestic dog can be informative for an improved understanding of human diseases. Indeed, this knowledge can also be directly beneficial for our canine companions. At the same time, human geneticists seek to solve the mystery of how and why we develop disease phenotypes. However, the knowledge gained in human studies might also help us understand the genetics behind canine disorders. This means that the adoption of the dog as an animal model to study genetic diseases has not always a univocal direction, thus strengthening the special relationship that we have built up with our best friends over a long time.

Acknowledgements

I should start saying that this is probably the part of the thesis most of the readers will directly jump to and read first…well, I've done it myself with every single thesis I have had in my hand, so I cannot blame whoever replicates my behavior. ☺

This work was carried out at the Department of Medical Biochemistry and Microbiology at Uppsala University, Uppsala, Sweden. I sincerely want to thank everyone I had the pleasure to work with and I already apologize in case someone will be unintentionally left out!

Gerli, I would like to express my deepest gratitude to you, not only for accepting me as a PhD student, but for all the patience, help, scientific support and guidance that you have given me during these years. You are a terrific scientist and I am really proud to have worked side by side with you. Thanks for always believing in me even when I was not doing it myself, and thanks for patiently leading me towards where I am now. You taught me how to focus on precise questions and how not to lose myself into the many different directions that practical work can take you. You taught me how to push myself in order to get scientifically better, and now I think I've achieved this goal. You taught me how to be less impatient and less impulsive. You taught me how to be less afraid of big challenges and how to be more self-confident…I think I am on my way. You have also given me an incredible personal support that only words cannot explain I guess. There have been complicated times, but you have always carefully listened to me and had a nice and positive word in every occasion. I really think I couldn't have had a better supervisor than you!

Kerstin, I am incredibly proud of having you as my co-supervisor. Working with you has been really inspirational, extremely challenging and motivating at the same time. Your attitude towards science is just fascinating: after talking to you I've always felt more determined to pursue my goals, even though they might have looked impossible to realize. Thanks for always supporting me!

Göran, thanks for your presence, availability and optimism. These are such important qualities!

Åke and Olle, you are such experts in your respective field and I feel lucky to collaborate with you. Your contribution has been just vital for the projects' start and continuation. On top of that, you always have really interesting and clever comments while looking at the big picture.

Cecilia, thanks for taking care of the whole dog group during my PhD early times. You have been an important person for all of us.

Manfred, I would like to thank you for all the patience you have had with me when I was lately knocking on your door and mail-bombing you! Many thanks for all the inspirational discussions we've had. Your passion and your excitement towards science are just contagious!

My deep thanks also go to everyone in the dog and human genomics groups. It has been really stimulating working with you! Daniel, for being a great collaborator and for the several scientific discussions we had; it has been a real pleasure working in direct contact with you! Jennifer, for being really cheerful, for sharing problems and solutions, for always coming up with new ideas and for always being critical when necessary, available and understanding when needed. Thank you also for taking care of the group lately! Marcin, for the interesting discussions about statistics and about the "P-value's value", and for always giving me smart analytical suggestions! Jessika, Iris (I am not done with you!), Lina, Fabiana, Erik, for being such important, skillful and valuable collaborators! I am really thankful for all your help, support and for all the discussions we have had. You have really played a big role during this time! Sergey, for the time you spent in trying to teach me immunology and for the free discussions about my projects.

Many thanks to all my coauthors for optimal and really fruitful collaborations…I think this is one of the secrets of good science.

I would also like to thank my office mates throughout the years for their help, patience and understanding. Probably too many times I yelled at my computer screen whenever something was not working…sorry for that, this might have seriously annoyed you.

Many thanks to Sharda, Maja[2], Vikki, Ginger, Nika, Leif, Calle (let me know if you need some more water from the aquarium! ☺), Ronnie, Nima, Jonas, Matt, Anna, Matts, Patrick, Andreas, Susanne, Daniél, Shumaila, Sangeet, Doreen, Freyja, Örjan, Mette, Simon, Emma, Mia and to all the other persons, present and past, at the Genomics corridor I didn't mention here…honestly, you know you have been somehow really instrumental either for my projects realization by giving me useful tips and suggestions, or for a good quality daily life at work. I am really proud of being a member of such high quality research corridor! I didn't realize it at first, but here we have written some important genetics research chapters! Eva, Åsa, Ulla and Jessica…thanks for the excellent support and help whenever a wet-lab related question or issue was coming up.

Thanks to Kais and Eva for being the best "supervisors" when the teaching was approaching! And many thanks to all the teaching lab assistants I've collaborated with throughout these years. It was fun working with you guys!

Many thanks to Alexis, Malin, Rehné, Veronica, Daniel and all the current and past administrative staff at IMBIM for helping me in many diverse matters.

I would like to thank Sofia, Tomas, Kumar, Susanne and all the other former colleagues at SLU for having represented a significant basis of my experience at Uppsala University.

I would also like to thank all the veterinarians and physicians, all the dogs and all the patients, which made all this work actually possible.

Miguel and Alvaro, you are among the brightest "junior" scientists I've known; many thanks for all our scientific discussions and for all the tips you have given me on the way. I am glad you are as much into football as me…it has been really fun to spend some time drinking beer and watching football together. And thanks for being not only colleagues, but also good friends who you can share your life stories, thoughts and concerns with.

Iris, you have also been one of such persons. Besides being a great colleague, you are the person I've had philosophical and endless talks with more than once, and the person always being really sympathetic and generous with advice and understanding on a countless number of situations. Sometimes we might look like slightly drifting apart for various reasons, but remember that friendship does not easily fade away! Ah…and take care of my German friend! ☺

Now I think it's also time to thank everyone has been part of my life outside work, and actually this part has meant a lot for me during this time in Uppsala.

It's really hard to explain in words what "I Linnei" have represented for me during these years…Bepi, Gigi, Ric, you have just been my second family and this I guess already explains everything. Writing the complete list of things we have done and we have shared, all the problems we have solved together and all the enjoyable moments we have lived, would be just redundant. You know what links us… if I had to talk in scientific terms, I would say that what binds us is like a covalent bond, really strong indeed. Thank you guys for being really good friends! Valentina, our friendship goes back in time…thank you especially for your help in the early times, for sharing some great salsa moments and for always being there. Dieghino and Fra' (and Niccol…a), thanks for being the good friends you are…I'm really glad to have you in my life. And try to visit me together with "the bomber", as you used to do…we can have some coffee and discuss science (Fra', now you know it!). Claire and Paola, you are my favorite sisters…thanks for the nice moments we've had, especially those under the Italian snow! Many, many thanks also go to Toni, Arianna, Valeria and Lorenzino, Albertino, Francesca[2], Moreno, Alessandro, Marco, Fulvio, Nicoló…thanks to all of you for your good friendship and for making me feel at home, even though I am quite far from it. Thanks also to my friends Pedro and Anna for the nice times we have spent together. Thanks to my poker mates…with you guys Mondays are actually meaningful and worth something! We have spent some really high-quality time together! Thanks Johan, for showing me that a Swed can actually be a bit Italian!

I would also like to thank Aly and Fila, Above and Beyond, Armin van Buuren and their mates for having accompanied me with their music throughout these years. Your contribution has been just, how to say, “uplifting”!

Thanks to all my friends who I meet every time I go back to Cantiano, my village in Italy. The time we spend together is always too little in terms of days, but I would say it is always pretty intense!

Thank you Ceci for being such an unexpected (and continuing) surprise!

Last but not the least…my family…grazie a tutti!

Grazie mamma e babbo per avermi sempre supportato, qualunque decisione io abbia preso. Grazie per essere sempre presenti, anche se piuttosto lontani. Grazie per tutti i consigli dati e per tutti i “cicchetti” che (ancora) mi date. Grazie perché non deve essere facile vedermi lontano per cosí tanto tempo, ma vi assicuro che ve la state cavando bene! Grazie nonna Elsa per essere semplicemente quello che sei sempre stata e sei tuttora. Grazie a tutti, zii, cugini e a chi purtroppo non c’é piú, ma che sicuramente mi é sempre stato vicino durante questi anni e continua a farlo.

References

1. Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, et al. Resolving the complexity of the human genome using single-molecule sequencing. Nature. 2015;517(7536):608-11. doi: 10.1038/nature13907.
2. Consortium UK, Walter K, Min JL, Huang J, Crooks L, Memari Y, et al. The UK10K project identifies rare variants in health and disease. Nature. 2015;526(7571):82-90. doi: 10.1038/nature14962.
3. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323(5910):133-8. doi: 10.1126/science.1162986.
4. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65. doi: 10.1038/nature11632.
5. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat Rev Genet. 2013;14(9):661-73. doi: 10.1038/nrg3502.
6. van Steenbeek FG, Hytonen MK, Leegwater PA, Lohi H. The canine era: the rise of a biomedical model. Anim Genet. 2016;47(5):519-27. doi: 10.1111/age.12460.
7. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78. doi: 10.1038/nature05911.
8. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, et al. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. PLoS Genet. 2011;7(10):e1002316. doi: 10.1371/journal.pgen.1002316.
9. The Genomes Project C. A global reference for human genetic variation. Nature. 2015;526(7571):68-74. doi: 10.1038/nature15393
http://www.nature.com/nature/journal/v526/n7571/abs/nature15393.html - supplementary-information.
10. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006;7(2):85-97. doi: 10.1038/nrg1767.
11. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, et al. Mapping and sequencing of structural variation from eight human genomes. Nature. 2008;453(7191):56-64. doi: 10.1038/nature06862.
12. Ziegler A, Konig IR, Pahlke F. A Statistical Approach to Genetic Epidemiology: Concepts and Applications, 2nd Edition. Wiley. 2010.
13. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet. 2008;9(4):255-66. doi: 10.1038/nrg2322.
14. Blekhman R, Man O, Herrmann L, Boyko AR, Indap A, Kosiol C, et al. Natural selection on genes that underlie human disease susceptibility. Curr Biol. 2008;18(12):883-9. doi: 10.1016/j.cub.2008.04.074.

15. Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, et al. Natural selection on protein-coding genes in the human genome. Nature. 2005;437(7062):1153-7. doi: 10.1038/nature04240.
16. Reich DE, Lander ES. On the allelic spectrum of human disease. Trends Genet. 2001;17(9):502-10.
17. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. Cell. 2011;147(1):32-43. doi: 10.1016/j.cell.2011.09.008.
18. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum Genet. 2013;132(10):1077-130. doi: 10.1007/s00439-013-1331-2.
19. Mitchell KJ. What is complex about complex disorders? Genome Biol. 2012;13(1):237. doi: 10.1186/gb-2012-13-1-237.
20. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 2008;40(6):695-701. doi: 10.1038/ng.f.136.
21. Fraser HB. Gene expression drives local adaptation in humans. Genome Res. 2013;23(7):1089-96. doi: 10.1101/gr.152710.112.
22. Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, Yen A, et al. Identifying recent adaptations in large-scale genomic data. Cell. 2013;152(4):703-13. doi: 10.1016/j.cell.2013.01.035.
23. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53. doi: 10.1038/nature08494.
24. Gibson G. Rare and common variants: twenty arguments. Nat Rev Genet. 2012;13(2):135-45. doi: 10.1038/nrg3118.
25. Lander ES. The new genomics: global views of biology. Science. 1996;274(5287):536-9.
26. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012;90(1):7-24. doi: 10.1016/j.ajhg.2011.11.029.
27. Karlsson EK, Sigurdsson S, Ivansson E, Thomas R, Elvers I, Wright J, et al. Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. Genome Biol. 2013;14(12):R132. doi: 10.1186/gb-2013-14-12-r132.
28. Seppala EH, Jokinen TS, Fukata M, Fukata Y, Webster MT, Karlsson EK, et al. LGI2 truncation causes a remitting focal epilepsy in dogs. PLoS Genet. 2011;7(7):e1002194. doi: 10.1371/journal.pgen.1002194.
29. Tang R, Noh HJ, Wang D, Sigurdsson S, Swofford R, Perloski M, et al. Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. Genome Biol. 2014;15(3):R25. doi: 10.1186/gb-2014-15-3-r25.
30. Karlsson EK, Lindblad-Toh K. Leader of the pack: gene mapping in dogs and other model organisms. Nat Rev Genet. 2008;9(9):713-25. doi: 10.1038/nrg2382.
31. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005;438(7069):803-19. doi: 10.1038/nature04338.
32. Sutter NB, Ostrander EA. Dog star rising: the canine genetic system. Nat Rev Genet. 2004;5(12):900-10. doi: 10.1038/nrg1492.

33. Wayne RK. Limb morphology of domestic and wild canids: the influence of development on morphologic change. J Morphol. 1986;187(3):301-19. doi: 10.1002/jmor.1051870304.
34. Ostrander EA, Kruglyak L. Unleashing the canine genome. Genome Res. 2000;10(9):1271-4.
35. Club AK. The complete dog book: 20th edition. New York: Ballantine Books. 2006.
36. Parker HG. Genomic analyses of modern dog breeds. Mamm Genome. 2012;23(1-2):19-27. doi: 10.1007/s00335-011-9387-6.
37. Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature. 2013;495(7441):360-4. doi: 10.1038/nature11837.
38. Coppinger R, Coppinger L. Dogs: a Startling New Understanding of Canine Origin, Behaviour and Evolution: Scribner; 2001.
39. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, et al. Linkage disequilibrium in the human genome. Nature. 2001;411(6834):199-204. doi: 10.1038/35075590.
40. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, et al. Extensive and breed-specific linkage disequilibrium in Canis familiaris. Genome Res. 2004;14(12):2388-96. doi: 10.1101/gr.3147604.
41. Patterson DF, Haskins ME, Jezyk PF, Giger U, Meyers-Wallen VN, Aguirre G, et al. Research on genetic diseases: reciprocal benefits to animals and man. J Am Vet Med Assoc. 1988;193(9):1131-44.
42. Sargan DR. IDID: inherited diseases in dogs: web-based information for canine inherited disease genetics. Mamm Genome. 2004;15(6):503-6. doi: 10.1007/s00335-004-3047-z.
43. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, et al. A polymorphic DNA marker genetically linked to Huntington's disease. Nature. 1983;306(5940):234-8.
44. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, et al. Identification of the cystic fibrosis gene: genetic analysis. Science. 1989;245(4922):1073-80.
45. Lin L, Faraco J, Li R, Kadotani H, Rogers W, Lin X, et al. The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. Cell. 1999;98(3):365-76.
46. Yuzbasiyan-Gurkan V, Blanton SH, Cao Y, Ferguson P, Li J, Venta PJ, et al. Linkage of a microsatellite marker to the canine copper toxicosis locus in Bedlington terriers. Am J Vet Res. 1997;58(1):23-7.
47. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005;6(2):95-108. doi: 10.1038/nrg1521.
48. Jorde LB. Linkage disequilibrium and the search for complex disease genes. Genome Res. 2000;10(10):1435-44.
49. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996;273(5281):1516-7.
50. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet. 2002;3(5):391-7. doi: 10.1038/nrg796.

51. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet. 2004;75(2):330-7. doi: 10.1086/422827.
52. Gaffney PM, Kearns GM, Shark KB, Ortmann WA, Selby SA, Malmgren ML, et al. A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. Proc Natl Acad Sci U S A. 1998;95(25):14875-9.
53. Nistico L, Buzzetti R, Pritchard LE, Van der Auwera B, Giovannini C, Bosi E, et al. The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. Hum Mol Genet. 1996;5(7):1075-80.
54. Rich SS, Weitkamp LR, Barbosa J. Genetic heterogeneity of insulin-dependent (type I) diabetes mellitus: evidence from a study of extended haplotypes. Am J Hum Genet. 1984;36(5):1015-23.
55. Velaga MR, Wilson V, Jennings CE, Owen CJ, Herington S, Donaldson PT, et al. The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. J Clin Endocrinol Metab. 2004;89(11):5862-5. doi: 10.1210/jc.2004-1108.
56. The International HapMap Project. Nature. 2003;426(6968):789-96. doi: 10.1038/nature02168.
57. A haplotype map of the human genome. Nature. 2005;437(7063):1299-320. doi: 10.1038/nature04226.
58. Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, et al. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science. 2009;324(5926):528-32. doi: 10.1126/science.1167936.
59. Cui Y, Li G, Li S, Wu R. Designs for linkage analysis and association studies of complex diseases. Methods Mol Biol. 2010;620:219-42. doi: 10.1007/978-1-60761-580-4_6.
60. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet. 2010;11(7):459-63. doi: 10.1038/nrg2813.
61. Voight BF, Pritchard JK. Confounding from cryptic relatedness in case-control association studies. PLoS Genet. 2005;1(3):e32. doi: 10.1371/journal.pgen.0010032.
62. Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA. 2008;299(11):1335-44. doi: 10.1001/jama.299.11.1335.
63. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, et al. Efficient control of population structure in model organism association mapping. Genetics. 2008;178(3):1709-23. doi: 10.1534/genetics.107.080101.
64. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. Nat Genet. 2010;42(4):355-60. doi: 10.1038/ng.546.
65. International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9. doi: 10.1038/nature10251.
66. Pagnacco G. Genetica Animale Applicata. Casa Editrice Ambrosiana. 2004.
67. Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. Genetics. 2011;187(2):367-83. doi: 10.1534/genetics.110.120907.

68. Bannasch D, Young A, Myers J, Truve K, Dickinson P, Gregg J, et al. Localization of canine brachycephaly using an across breed mapping approach. PLoS One. 2010;5(3):e9632. doi: 10.1371/journal.pone.0009632.
69. Olsson M, Meadows JR, Truve K, Rosengren Pielberg G, Puppo F, Mauceli E, et al. A novel unstable duplication upstream of HAS2 predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs. PLoS Genet. 2011;7(3):e1001332. doi: 10.1371/journal.pgen.1001332.
70. Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet. 2013;14(6):379-89. doi: 10.1038/nrg3472.
71. Thompson JR, Attia J, Minelli C. The meta-analysis of genome-wide association studies. Brief Bioinform. 2011;12(3):259-69. doi: 10.1093/bib/bbr020.
72. Zeggini E, Ioannidis JP. Meta-analysis in genome-wide association studies. Pharmacogenomics. 2009;10(2):191-201. doi: 10.2217/14622416.10.2.191.
73. Grada A, Weinbrecht K. Next-generation sequencing: methodology and application. J Invest Dermatol. 2013;133(8):e11. doi: 10.1038/jid.2013.248.
74. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. Cell. 2013;155(1):27-38. doi: 10.1016/j.cell.2013.09.006.
75. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921. doi: 10.1038/35057062.
76. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30(9):418-26. doi: 10.1016/j.tig.2014.07.001.
77. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet. 2008;24(3):133-41. doi: 10.1016/j.tig.2007.12.007.
78. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. J Genet Genomics. 2011;38(3):95-109. doi: 10.1016/j.jgg.2011.02.003.
79. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol. 2009;4(4):265-70. doi: 10.1038/nnano.2009.12.
80. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11(1):31-46. doi: 10.1038/nrg2626.
81. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. Nat Methods. 2010;7(2):111-8. doi: 10.1038/nmeth.1419.
82. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010;42(1):30-5. doi: 10.1038/ng.499.
83. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009;27(2):182-9. doi: 10.1038/nbt.1523.
84. Grover CE, Salmon A, Wendel JF. Targeted sequence capture as a powerful tool for evolutionary analysis. Am J Bot. 2012;99(2):312-9. doi: 10.3732/ajb.1100323.
85. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010;11(7):499-511. doi: 10.1038/nrg2796.

86. Boomsma DI, Wijmenga C, Slagboom EP, Swertz MA, Karssen LC, Abdellaoui A, et al. The Genome of the Netherlands: design, and project goals. Eur J Hum Genet. 2014;22(2):221-7. doi: 10.1038/ejhg.2013.118.
87. Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A, et al. Large-scale whole-genome sequencing of the Icelandic population. Nat Genet. 2015;47(5):435-44. doi: 10.1038/ng.3247.
88. van Leeuwen EM, Karssen LC, Deelen J, Isaacs A, Medina-Gomez C, Mbarek H, et al. Genome of The Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. Nat Commun. 2015;6:6065. doi: 10.1038/ncomms7065.
89. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011;13(1):101. doi: 10.1186/ar3204.
90. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet. 2011;43(12):1193-201. doi: 10.1038/ng.998.
91. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet. 2012;8(8):e1002793. doi: 10.1371/journal.pgen.1002793.
92. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010;28(10):1045-8. doi: 10.1038/nbt1010-1045.
93. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74. doi: 10.1038/nature11247.
94. Lowe WL, Jr., Reddy TE. Genomic approaches for understanding the genetics of complex disease. Genome Res. 2015;25(10):1432-41. doi: 10.1101/gr.190603.115.
95. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol. 2015;16:22. doi: 10.1186/s13059-014-0560-6.
96. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5. doi: 10.1038/ng.2653.
97. McGonagle D, McDermott MF. A proposed classification of the immunological diseases. PLoS Med. 2006;3(8):e297. doi: 10.1371/journal.pmed.0030297.
98. Abbas AK, Lichtman AH, Pillai S. Cellular and molecular immunology (7th edition). Philadelphia: Elsevier Saunders. 2012.
99. Cooper GS, Bynum ML, Somers EC. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. J Autoimmun. 2009;33(3-4):197-207. doi: 10.1016/j.jaut.2009.09.008.
100. Whitacre CC, Reingold SC, O'Looney PA. A gender gap in autoimmunity. Science. 1999;283(5406):1277-8.
101. Liu K, Kurien BT, Zimmerman SL, Kaufman KM, Taft DH, Kottyan LC, et al. X Chromosome Dose and Sex Bias in Autoimmune Diseases: Increased Prevalence of 47,XXX in Systemic Lupus Erythematosus and Sjogren's Syndrome. Arthritis Rheumatol. 2016;68(5):1290-300. doi: 10.1002/art.39560.
102. Miech RP. The role of fetal microchimerism in autoimmune disease. Int J Clin Exp Med. 2010;3(2):164-8.

103. Quintero OL, Amador-Patarroyo MJ, Montoya-Ortiz G, Rojas-Villarraga A, Anaya JM. Autoimmune disease and gender: plausible mechanisms for the female predominance of autoimmunity. J Autoimmun. 2012;38(2-3):J109-19. doi: 10.1016/j.jaut.2011.10.003.
104. Cantorna MT, Mahon BD. Mounting evidence for vitamin D as an environmental factor affecting autoimmune disease prevalence. Exp Biol Med (Maywood). 2004;229(11):1136-42.
105. Costenbader KH, Gay S, Alarcon-Riquelme ME, Iaccarino L, Doria A. Genes, epigenetic regulation and environmental factors: which is the most relevant in developing autoimmune diseases? Autoimmun Rev. 2012;11(8):604-9. doi: 10.1016/j.autrev.2011.10.022.
106. Effraimidis G, Wiersinga WM. Mechanisms in endocrinology: autoimmune thyroid disease: old and new players. Eur J Endocrinol. 2014;170(6):R241-52. doi: 10.1530/EJE-14-0047.
107. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015;518(7539):337-43. doi: 10.1038/nature13835.
108. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nat Rev Genet. 2009;10(1):43-55. doi: 10.1038/nrg2489.
109. Zhernakova A, Withoff S, Wijmenga C. Clinical implications of shared genetics and pathogenesis in autoimmune diseases. Nat Rev Endocrinol. 2013;9(11):646-59. doi: 10.1038/nrendo.2013.161.
110. Ricano-Ponce I, Wijmenga C. Mapping of immune-mediated disease genes. Annu Rev Genomics Hum Genet. 2013;14:325-53. doi: 10.1146/annurev-genom-091212-153450.
111. Ivansson EL, Megquier K, Kozyrev SV, Muren E, Korberg IB, Swofford R, et al. Variants within the SP110 nuclear body protein modify risk of canine degenerative myelopathy. Proc Natl Acad Sci U S A. 2016;113(22):E3091-100. doi: 10.1073/pnas.1600084113.
112. Truve K, Dickinson P, Xiong A, York D, Jayashankar K, Pielberg G, et al. Utilizing the Dog Genome in the Search for Novel Candidate Genes Involved in Glioma Development-Genome Wide Association Mapping followed by Targeted Massive Parallel Sequencing Identifies a Strongly Associated Locus. PLoS Genet. 2016;12(5):e1006000. doi: 10.1371/journal.pgen.1006000.
113. Wilbe M, Jokinen P, Truve K, Seppala EH, Karlsson EK, Biagi T, et al. Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. Nat Genet. 2010;42(3):250-4. doi: 10.1038/ng.525.
114. Gaitonde DY, Rowley KD, Sweeney LB. Hypothyroidism: an update. Am Fam Physician. 2012;86(3):244-51.
115. Khandelwal D, Tandon N. Overt and subclinical hypothyroidism: who to treat and how. Drugs. 2012;72(1):17-33. doi: 10.2165/11598070-000000000-00000.
116. Roberts CG, Ladenson PW. Hypothyroidism. Lancet. 2004;363(9411):793-803. doi: 10.1016/S0140-6736(04)15696-1.
117. Laurberg P, Cerqueira C, Ovesen L, Rasmussen LB, Perrild H, Andersen S, et al. Iodine intake as a determinant of thyroid disorders in populations. Best Pract Res Clin Endocrinol Metab. 2010;24(1):13-27. doi: 10.1016/j.beem.2009.08.013.
118. Jacobson DL, Gange SJ, Rose NR, Graham NM. Epidemiology and estimated population burden of selected autoimmune diseases in the United States. Clin Immunol Immunopathol. 1997;84(3):223-43.

119. Fairweather D, Rose NR. Coxsackievirus-induced myocarditis in mice: a model of autoimmune disease for studying immunotoxicity. Methods. 2007;41(1):118-22. doi: 10.1016/j.ymeth.2006.07.009.
120. Vanderpump MP. The epidemiology of thyroid disease. Br Med Bull. 2011;99:39-51. doi: 10.1093/bmb/ldr030.
121. Wang C, Crapo LM. The epidemiology of thyroid disease and implications for screening. Endocrinol Metab Clin North Am. 1997;26(1):189-218.
122. Pearce EN, Farwell AP, Braverman LE. Thyroiditis. N Engl J Med. 2003;348(26):2646-55. doi: 10.1056/NEJMra021194.
123. Prummel MF, Strieder T, Wiersinga WM. The environment and autoimmune thyroid diseases. Eur J Endocrinol. 2004;150(5):605-18.
124. Garber JR, Cobin RH, Gharib H, Hennessey JV, Klein I, Mechanick JI, et al. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. Endocr Pract. 2012;18(6):988-1028. doi: 10.4158/EP12280.GL.
125. Brix TH, Kyvik KO, Hegedus L. A population-based study of chronic autoimmune hypothyroidism in Danish twins. J Clin Endocrinol Metab. 2000;85(2):536-9. doi: 10.1210/jcem.85.2.6385.
126. Dittmar M, Libich C, Brenzel T, Kahaly GJ. Increased familial clustering of autoimmune thyroid diseases. Horm Metab Res. 2011;43(3):200-4. doi: 10.1055/s-0031-1271619.
127. Hansen PS, Brix TH, Iachine I, Kyvik KO, Hegedus L. The relative importance of genetic and environmental effects for the early stages of thyroid autoimmunity: a study of healthy Danish twins. Eur J Endocrinol. 2006;154(1):29-38. doi: 10.1530/eje.1.02060.
128. Villanueva R, Greenberg DA, Davies TF, Tomer Y. Sibling recurrence risk in autoimmune thyroid disease. Thyroid. 2003;13(8):761-4. doi: 10.1089/105072503768499653.
129. Tomer Y, Ban Y, Concepcion E, Barbesino G, Villanueva R, Greenberg DA, et al. Common and unique susceptibility loci in Graves and Hashimoto diseases: results of whole-genome screening in a data set of 102 multiplex families. Am J Hum Genet. 2003;73(4):736-47. doi: 10.1086/378588.
130. Colobran R, Armengol Mdel P, Faner R, Gartner M, Tykocinski LO, Lucas A, et al. Association of an SNP with intrathymic transcription of TSHR and Graves' disease: a role for defective thymic tolerance. Hum Mol Genet. 2011;20(17):3415-23. doi: 10.1093/hmg/ddr247.
131. Farid NR, Sampson L, Moens H, Barnard JM. The association of goitrous autoimmune thyroiditis with HLA-DR5. Tissue Antigens. 1981;17(3):265-8.
132. Moens H, Farid NR, Sampson L, Noel EP, Barnard JM. Hashimoto's thyroiditis is associated with HLA-DRw3. N Engl J Med. 1978;299(3):133-4. doi: 10.1056/NEJM197807202990306.
133. Petrone A, Giorgi G, Mesturino CA, Capizzi M, Cascino I, Nistico L, et al. Association of DRB1*04-DQB1*0301 haplotype and lack of association of two polymorphic sites at CTLA-4 gene with Hashimoto's thyroiditis in an Italian population. Thyroid. 2001;11(2):171-5. doi: 10.1089/105072501300042901.
134. Zamani M, Spaepen M, Bex M, Bouillon R, Cassiman JJ. Primary role of the HLA class II DRB1*0301 allele in Graves disease. Am J Med Genet. 2000;95(5):432-7.
135. Kotsa K, Watson PF, Weetman AP. A CTLA-4 gene polymorphism is associated with both Graves disease and autoimmune hypothyroidism. Clin Endocrinol (Oxf). 1997;46(5):551-4.

136. Criswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, et al. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. Am J Hum Genet. 2005;76(4):561-71. doi: 10.1086/429096.
137. Brand OJ, Lowe CE, Heward JM, Franklyn JA, Cooper JD, Todd JA, et al. Association of the interleukin-2 receptor alpha (IL-2Ralpha)/CD25 gene region with Graves' disease using a multilocus test and tag SNPs. Clin Endocrinol (Oxf). 2007;66(4):508-12. doi: 10.1111/j.1365-2265.2007.02762.x.
138. Ban Y, Tozaki T, Tobe T, Ban Y, Jacobson EM, Concepcion ES, et al. The regulatory T cell gene FOXP3 and genetic susceptibility to thyroid autoimmunity: an association analysis in Caucasian and Japanese cohorts. J Autoimmun. 2007;28(4):201-7. doi: 10.1016/j.jaut.2007.02.016.
139. Tomer Y, Concepcion E, Greenberg DA. A C/T single-nucleotide polymorphism in the region of the CD40 gene is associated with Graves' disease. Thyroid. 2002;12(12):1129-35. doi: 10.1089/105072502321085234.
140. Simmonds MJ, Brand OJ, Barrett JC, Newby PR, Franklyn JA, Gough SC. Association of Fc receptor-like 5 (FCRL5) with Graves' disease is secondary to the effect of FCRL3. Clin Endocrinol (Oxf). 2010;73(5):654-60. doi: 10.1111/j.1365-2265.2010.03843.x.
141. Wellcome Trust Case Control C, Australo-Anglo-American Spondylitis C, Burton PR, Clayton DG, Cardon LR, Craddock N, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet. 2007;39(11):1329-37. doi: 10.1038/ng.2007.17.
142. Cooper JD, Simmonds MJ, Walker NM, Burren O, Brand OJ, Guo H, et al. Seven newly identified loci for autoimmune thyroid disease. Hum Mol Genet. 2012;21(23):5202-8. doi: 10.1093/hmg/dds357.
143. Medici M, Porcu E, Pistis G, Teumer A, Brown SJ, Jensen RA, et al. Identification of novel genetic Loci associated with thyroid peroxidase antibodies and clinical thyroid disease. PLoS Genet. 2014;10(2):e1004123. doi: 10.1371/journal.pgen.1004123.
144. Oryoji D, Ueda S, Yamamoto K, Yoshimura Noh J, Okamura K, Noda M, et al. Identification of a Hashimoto thyroiditis susceptibility locus via a genome-wide comparison with Graves' disease. J Clin Endocrinol Metab. 2015;100(2):E319-24. doi: 10.1210/jc.2014-3431.
145. Ferguson DC. Testing for hypothyroidism in dogs. Vet Clin North Am Small Anim Pract. 2007;37(4):647-69, v. doi: 10.1016/j.cvsm.2007.05.015.
146. Mooney CT. Canine hypothyroidism: a review of aetiology and diagnosis. N Z Vet J. 2011;59(3):105-14. doi: 10.1080/00480169.2011.563729.
147. Dodgson SE, Day R, Fyfe JC. Congenital hypothyroidism with goiter in Tenterfield terriers. J Vet Intern Med. 2012;26(6):1350-7. doi: 10.1111/j.1939-1676.2012.01015.x.
148. Fyfe JC, Kampschmidt K, Dang V, Poteet BA, He Q, Lowrie C, et al. Congenital hypothyroidism with goiter in toy fox terriers. J Vet Intern Med. 2003;17(1):50-7.
149. Graham PA, Refsal KR, Nachreiner RF. Etiopathologic findings of canine hypothyroidism. Vet Clin North Am Small Anim Pract. 2007;37(4):617-31, v. doi: 10.1016/j.cvsm.2007.05.002.
150. Happ GM. Thyroiditis--a model canine autoimmune disease. Adv Vet Sci Comp Med. 1995;39:97-139.
151. Lucke VM, Gaskell CJ, Wotton PR. Thyroid pathology in canine hypothyroidism. J Comp Pathol. 1983;93(3):415-21.

152. Beierwaltes WH, Nishiyama RH. Dog thyroiditis: occurrence and similarity to Hashimoto's struma. Endocrinology. 1968;83(3):501-8. doi: 10.1210/endo-83-3-501.
153. Benjamin SA, Stephens LC, Hamilton BF, Saunders WJ, Lee AC, Angleton GM, et al. Associations between lymphocytic thyroiditis, hypothyroidism, and thyroid neoplasia in beagles. Vet Pathol. 1996;33(5):486-94.
154. Egenvall A, Bonnett BN, Olson P, Hedhammar A. Gender, age, breed and distribution of morbidity and mortality in insured dogs in Sweden during 1995 and 1996. Vet Rec. 2000;146(18):519-25.
155. Kennedy LJ, Huson HJ, Leonard J, Angles JM, Fox LE, Wojciechowski JW, et al. Association of hypothyroid disease in Doberman Pinscher dogs with a rare major histocompatibility complex DLA class II haplotype. Tissue Antigens. 2006;67(1):53-6. doi: 10.1111/j.1399-0039.2005.00518.x.
156. Nachreiner RF, Refsal KR, Graham PA, Bowman MM. Prevalence of serum thyroid hormone autoantibodies in dogs with clinical signs of hypothyroidism. J Am Vet Med Assoc. 2002;220(4):466-71.
157. Scott DW, Paradis M. A survey of canine and feline skin disorders seen in a university practice: Small Animal Clinic, University of Montreal, Saint-Hyacinthe, Quebec (1987-1988). Can Vet J. 1990;31(12):830-5.
158. Graham PA, Nachreiner RF, Refsal KR, Provencher-Bolliger AL. Lymphocytic thyroiditis. Vet Clin North Am Small Anim Pract. 2001;31(5):915-33, vi-vii.
159. Benjamin SA, Saunders WJ, Lee AC, Angleton GM, Stephens LC, Mallinckrodt CH. Non-neoplastic and neoplastic thyroid disease in beagles irradiated during prenatal and postnatal development. Radiat Res. 1997;147(4):422-30.
160. Ahlgren J, Uimari P. Heritability of hypothyroidism in the Finnish Hovawart population. Acta Vet Scand. 2016;58(1):39. doi: 10.1186/s13028-016-0221-8.
161. Kennedy LJ, Quarmby S, Happ GM, Barnes A, Ramsey IK, Dixon RM, et al. Association of canine hypothyroidism with a common major histocompatibility complex DLA class II allele. Tissue Antigens. 2006;68(1):82-6. doi: 10.1111/j.1399-0039.2006.00614.x.
162. Wilbe M, Sundberg K, Hansen IR, Strandberg E, Nachreiner RF, Hedhammar A, et al. Increased genetic risk or protection for canine autoimmune lymphocytic thyroiditis in Giant Schnauzers depends on DLA class II genotype. Tissue Antigens. 2010;75(6):712-9. doi: 10.1111/j.1399-0039.2010.01449.x.
163. Zelissen PM, Bast EJ, Croughs RJ. Associated autoimmunity in Addison's disease. J Autoimmun. 1995;8(1):121-30. doi: 10.1006/jaut.1995.0009.
164. Lovas K, Husebye ES. Addison's disease. Lancet. 2005;365(9476):2058-61. doi: 10.1016/S0140-6736(05)66700-1.
165. Husebye ES, Allolio B, Arlt W, Badenhoop K, Bensing S, Betterle C, et al. Consensus statement on the diagnosis, treatment and follow-up of patients with primary adrenal insufficiency. J Intern Med. 2014;275(2):104-15. doi: 10.1111/joim.12162.
166. Bjornsdottir S, Sundstrom A, Ludvigsson JF, Blomqvist P, Kampe O, Bensing S. Drug prescription patterns in patients with Addison's disease: a Swedish population-based cohort study. J Clin Endocrinol Metab. 2013;98(5):2009-18. doi: 10.1210/jc.2012-3561.

167. Erichsen MM, Lovas K, Skinningsrud B, Wolff AB, Undlien DE, Svartberg J, et al. Clinical, immunological, and genetic features of autoimmune primary adrenal insufficiency: observations from a Norwegian registry. J Clin Endocrinol Metab. 2009;94(12):4882-90. doi: 10.1210/jc.2009-1368.
168. Kong MF, Jeffcoate W. Eighty-six cases of Addison's disease. Clin Endocrinol (Oxf). 1994;41(6):757-61.
169. Laureti S, Vecchi L, Santeusanio F, Falorni A. Is the prevalence of Addison's disease underestimated? J Clin Endocrinol Metab. 1999;84(5):1762. doi: 10.1210/jcem.84.5.5677-7.
170. Lovas K, Husebye ES. High prevalence and increasing incidence of Addison's disease in western Norway. Clin Endocrinol (Oxf). 2002;56(6):787-91.
171. Mason AS, Meade TW, Lee JA, Morris JN. Epidemiological and clinical picture of Addison's disease. Lancet. 1968;2(7571):744-7.
172. Willis AC, Vince FP. The prevalence of Addison's disease in Coventry, UK. Postgrad Med J. 1997;73(859):286-8.
173. Falorni A, Laureti S, De Bellis A, Zanchetta R, Tiberti C, Arnaldi G, et al. Italian addison network study: update of diagnostic criteria for the etiological classification of primary adrenal insufficiency. J Clin Endocrinol Metab. 2004;89(4):1598-604. doi: 10.1210/jc.2003-030954.
174. Winqvist O, Karlsson FA, Kampe O. 21-Hydroxylase, a major autoantigen in idiopathic Addison's disease. Lancet. 1992;339(8809):1559-62.
175. Betterle C, Morlin L. Autoimmune Addison's disease. Endocr Dev. 2011;20:161-72. doi: 10.1159/000321239.
176. Falorni A, Bini V, Betterle C, Brozzetti A, Castano L, Fichna M, et al. Determination of 21-hydroxylase autoantibodies: inter-laboratory concordance in the Euradrenal International Serum Exchange Program. Clin Chem Lab Med. 2015;53(11):1761-70. doi: 10.1515/cclm-2014-1106.
177. Cutolo M. Autoimmune polyendocrine syndromes. Autoimmun Rev. 2014;13(2):85-9. doi: 10.1016/j.autrev.2013.07.006.
178. Betterle C, Volpato M, Greggio AN, Presotto F. Type 2 polyglandular autoimmune disease (Schmidt's syndrome). J Pediatr Endocrinol Metab. 1996;9 Suppl 1:113-23.
179. Mitchell AL, Pearce SH. Autoimmune Addison disease: pathophysiology and genetic complexity. Nat Rev Endocrinol. 2012;8(5):306-16. doi: 10.1038/nrendo.2011.245.
180. Fairchild RS, Schimke RN, Abdou NI. Immunoregulation abnormalities in familial Addison's disease. J Clin Endocrinol Metab. 1980;51(5):1074-7. doi: 10.1210/jcem-51-5-1074.
181. Heggarty H. Addison's disease in identical twins. Br Med J. 1968;1(5591):559.
182. Hewitt PH. Addison's disease occurring in sisters. Br Med J. 1957;2(5060):1530-1.
183. Mitchell AL, Boe Wolff A, MacArthur K, Weaver JU, Vaidya B, Swedish Addision Registry Study G, et al. Linkage Analysis in Autoimmune Addison's Disease: NFATC1 as a Potential Novel Susceptibility Locus. PLoS One. 2015;10(6):e0123550. doi: 10.1371/journal.pone.0123550.
184. Simmonds JP, Lister J. Auto-immune Addison's disease in identical twins. Postgrad Med J. 1978;54(634):552-4.
185. Smith ME, Gough J, Galpin OP. Addison's Disease in Identical Twins. Br Med J. 1963;2(5368):1316.

186. Mitchell AL, Macarthur KD, Gan EH, Baggott LE, Wolff AS, Skinningsrud B, et al. Association of autoimmune Addison's disease with alleles of STAT4 and GATA3 in European cohorts. PLoS One. 2014;9(3):e88991. doi: 10.1371/journal.pone.0088991.
187. Hanson JM, Tengvall K, Bonnett BN, Hedhammar A. Naturally Occurring Adrenocortical Insufficiency--An Epidemiological Study Based on a Swedish-Insured Dog Population of 525,028 Dogs. J Vet Intern Med. 2016;30(1):76-84. doi: 10.1111/jvim.13815.
188. Famula TR, Belanger JM, Oberbauer AM. Heritability and complex segregation analysis of hypoadrenocorticism in the standard poodle. J Small Anim Pract. 2003;44(1):8-12.
189. Hughes AM, Nelson RW, Famula TR, Bannasch DL. Clinical features and heritability of hypoadrenocorticism in Nova Scotia Duck Tolling Retrievers: 25 cases (1994-2006). J Am Vet Med Assoc. 2007;231(3):407-12. doi: 10.2460/javma.231.3.407.
190. Oberbauer AM, Bell JS, Belanger JM, Famula TR. Genetic evaluation of Addison's disease in the Portuguese Water Dog. BMC Vet Res. 2006;2:15. doi: 10.1186/1746-6148-2-15.
191. Chase K, Sargan D, Miller K, Ostrander EA, Lark KG. Understanding the genetics of autoimmune disease: two loci that regulate late onset Addison's disease in Portuguese Water Dogs. Int J Immunogenet. 2006;33(3):179-84. doi: 10.1111/j.1744-313X.2006.00593.x.
192. Hughes AM, Jokinen P, Bannasch DL, Lohi H, Oberbauer AM. Association of a dog leukocyte antigen class II haplotype with hypoadrenocorticism in Nova Scotia Duck Tolling Retrievers. Tissue Antigens. 2010;75(6):684-90. doi: 10.1111/j.1399-0039.2010.01440.x.
193. Bowen D, Schaer M, Riley W. Autoimmune polyglandular syndrome in a dog: a case report. sep-oct1986;v. 22.
194. Friedenberg SG, Lunn KF, Meurs KM. Evaluation of the genetic basis of primary hypoadrenocorticism in Standard Poodles using SNP array genotyping and whole-genome sequencing. Mamm Genome. 2017;28(1-2):56-65. doi: 10.1007/s00335-016-9671-6.
195. Panciera DL. Hypothyroidism in dogs: 66 cases (1987-1992). J Am Vet Med Assoc. 1994;204(5):761-7.
196. Ferm K, Bjornerfeldt S, Karlsson A, Andersson G, Nachreiner R, Hedhammar A. Prevalence of diagnostic characteristics indicating canine autoimmune lymphocytic thyroiditis in giant schnauzer and hovawart dogs. J Small Anim Pract. 2009;50(4):176-9. doi: 10.1111/j.1748-5827.2008.00696.x.
197. Ziener ML, Dahlgren S, Thoresen SI, Lingaas F. Genetics and epidemiology of hypothyroidism and symmetrical onychomadesis in the Gordon setter and the English setter. Canine Genet Epidemiol. 2015;2:12. doi: 10.1186/s40575-015-0025-6.
198. Chakera AJ, Pearce SH, Vaidya B. Treatment for primary hypothyroidism: current approaches and future possibilities. Drug Des Devel Ther. 2012;6:1-11. doi: 10.2147/DDDT.S12894.
199. Milne KL, Hayes HM, Jr. Epidemiologic features of canine hypothyroidism. Cornell Vet. 1981;71(1):3-14.
200. Boretti FS, Reusch CE. Endogenous TSH in the diagnosis of hypothyroidism in dogs. Schweiz Arch Tierheilkd. 2004;146(4):183-8. doi: 10.1024/0036-7281.146.4.183.

201. Dixon RM, Mooney CT. Evaluation of serum free thyroxine and thyrotropin concentrations in the diagnosis of canine hypothyroidism. J Small Anim Pract. 1999;40(2):72-8.
202. Kantrowitz LB, Peterson ME, Melian C, Nichols R. Serum total thyroxine, total triiodothyronine, free thyroxine, and thyrotropin concentrations in dogs with nonthyroidal disease. J Am Vet Med Assoc. 2001;219(6):765-9.
203. Paradis M, Sauve F, Charest J, Refsal KR, Moreau M, Dupuis J. Effects of moderate to severe osteoarthritis on canine thyroid function. Can Vet J. 2003;44(5):407-12.
204. Daminet S, Ferguson DC. Influence of drugs on thyroid function in dogs. J Vet Intern Med. 2003;17(4):463-72.
205. Ramsey IK, Evans H, Herrtage ME. Thyroid-stimulating hormone and total thyroxine concentrations in euthyroid, sick euthyroid and hypothyroid dogs. J Small Anim Pract. 1997;38(12):540-5.
206. Dixon RM, Mooney CT. Canine serum thyroglobulin autoantibodies in health, hypothyroidism and non-thyroidal illness. Res Vet Sci. 1999;66(3):243-6. doi: 10.1053/rvsc.1998.0268.
207. Kennedy LJ, Barnes A, Short A, Brown JJ, Seddon J, Fleeman L, et al. Canine DLA diversity: 3. Disease studies. Tissue Antigens. 2007;69 Suppl 1:292-6. doi: 10.1111/j.1399-0039.2006.00781.x.
208. Meurs KM, Mauceli E, Lahmers S, Acland GM, White SN, Lindblad-Toh K. Genome-wide association identifies a deletion in the 3' untranslated region of striatin in a canine model of arrhythmogenic right ventricular cardiomyopathy. Hum Genet. 2010;128(3):315-24. doi: 10.1007/s00439-010-0855-y.
209. Tengvall K, Kierczak M, Bergvall K, Olsson M, Frankowiack M, Farias FH, et al. Genome-wide analysis in German shepherd dogs reveals association of a locus on CFA 27 with atopic dermatitis. PLoS Genet. 2013;9(5):e1003475. doi: 10.1371/journal.pgen.1003475.
210. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, et al. Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet. 2007;39(11):1321-8. doi: 10.1038/ng.2007.10.
211. Kalay E, Li Y, Uzumcu A, Uyguner O, Collin RW, Caylan R, et al. Mutations in the lipoma HMGIC fusion partner-like 5 (LHFPL5) gene cause autosomal recessive nonsyndromic hearing loss. Hum Mutat. 2006;27(7):633-9. doi: 10.1002/humu.20368.
212. Shabbir MI, Ahmed ZM, Khan SY, Riazuddin S, Waryah AM, Khan SN, et al. Mutations of human TMHS cause recessively inherited non-syndromic hearing loss. J Med Genet. 2006;43(8):634-40. doi: 10.1136/jmg.2005.039834.
213. Longo-Guess CM, Gagnon LH, Cook SA, Wu J, Zheng QY, Johnson KR. A missense mutation in the previously undescribed gene Tmhs underlies deafness in hurry-scurry (hscy) mice. Proc Natl Acad Sci U S A. 2005;102(22):7894-9. doi: 10.1073/pnas.0500760102.
214. Toker A, Chin YR. Akt-ing up on SRPK1: oncogene or tumor suppressor? Mol Cell. 2014;54(3):329-30. doi: 10.1016/j.molcel.2014.04.020.
215. Nousiainen L, Sillanpaa M, Jiang M, Thompson J, Taipale J, Julkunen I. Human kinome analysis reveals novel kinases contributing to virus infection and retinoic-acid inducible gene I-induced type I and type III IFN gene expression. Innate Immun. 2013;19(5):516-30. doi: 10.1177/1753425912473345.

216. Prescott EL, Brimacombe CL, Hartley M, Bell I, Graham S, Roberts S. Human papillomavirus type 1 E1^E4 protein is a potent inhibitor of the serine-arginine (SR) protein kinase SRPK1 and inhibits phosphorylation of host SR proteins and of the viral transcription and replication regulator E2. J Virol. 2014;88(21):12599-611. doi: 10.1128/JVI.02029-14.
217. Kamachi M, Le TM, Kim SJ, Geiger ME, Anderson P, Utz PJ. Human autoimmune sera as molecular probes for the identification of an autoantigen kinase signaling pathway. J Exp Med. 2002;196(9):1213-25.
218. Dror AA, Lenz DR, Shivatzki S, Cohen K, Ashur-Fabian O, Avraham KB. Atrophic thyroid follicles and inner ear defects reminiscent of cochlear hypothyroidism in Slc26a4-related deafness. Mamm Genome. 2014;25(7-8):304-16. doi: 10.1007/s00335-014-9515-1.
219. Mavragani CP, Niewold TB, Chatzigeorgiou A, Danielides S, Thomas D, Kirou KA, et al. Increased serum type I interferon activity in organ-specific autoimmune disorders: clinical, imaging, and serological associations. Front Immunol. 2013;4:238. doi: 10.3389/fimmu.2013.00238.
220. Bauer JW, Baechler EC, Petri M, Batliwalla FM, Crawford D, Ortmann WA, et al. Elevated serum levels of interferon-regulated chemokines are biomarkers for active human systemic lupus erythematosus. PLoS Med. 2006;3(12):e491. doi: 10.1371/journal.pmed.0030491.
221. Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J Exp Med. 2003;197(6):711-23. doi: 10.1084/jem.20021553.
222. Bronson PG, Chaivorapol C, Ortmann W, Behrens TW, Graham RR. The genetics of type I interferon in systemic lupus erythematosus. Curr Opin Immunol. 2012;24(5):530-7. doi: 10.1016/j.coi.2012.07.008.
223. Dall'era MC, Cardarelli PM, Preston BT, Witte A, Davis JC, Jr. Type I interferon correlates with serological and clinical manifestations of SLE. Ann Rheum Dis. 2005;64(12):1692-7. doi: 10.1136/ard.2004.033753.
224. Custro N, Montalto G, Scafidi V, Soresi M, Gallo S, Tripi S, et al. Prospective study on thyroid autoimmunity and dysfunction related to chronic hepatitis C and interferon therapy. J Endocrinol Invest. 1997;20(7):374-80. doi: 10.1007/BF03347987.
225. Nagayama Y, Ohta K, Tsuruta M, Takeshita A, Kimura H, Hamasaki K, et al. Exacerbation of thyroid autoimmunity by interferon alpha treatment in patients with chronic viral hepatitis: our studies and review of the literature. Endocr J. 1994;41(5):565-72.
226. Hoeppner MP, Lundquist A, Pirun M, Meadows JR, Zamani N, Johnson J, et al. An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. PLoS One. 2014;9(3):e91172. doi: 10.1371/journal.pone.0091172.
227. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A. 2009;106(45):19096-101. doi: 10.1073/pnas.0910672106.
228. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009;461(7261):272-6. doi: 10.1038/nature08250.
229. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution map of human evolutionary constraint using 29 mammals. Nature. 2011;478(7370):476-82. doi: 10.1038/nature10530.

230. Garner C, Ahn R, Ding YC, Steele L, Stoven S, Green PH, et al. Genome-wide association study of celiac disease in North America confirms FRMD4B as new celiac locus. PLoS One. 2014;9(7):e101428. doi: 10.1371/journal.pone.0101428.
231. Grant SF, Qu HQ, Bradfield JP, Marchand L, Kim CE, Glessner JT, et al. Follow-up analysis of genome-wide association data identifies novel loci for type 1 diabetes. Diabetes. 2009;58(1):290-5. doi: 10.2337/db08-1022.
232. Jin Y, Birlea SA, Fain PR, Ferrara TM, Ben S, Riccardi SL, et al. Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. Nat Genet. 2012;44(6):676-80. doi: 10.1038/ng.2272.
233. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119-24. doi: 10.1038/nature11582.
234. Gombos Z, Hermann R, Kiviniemi M, Nejentsev S, Reimand K, Fadeyev V, et al. Analysis of extended human leukocyte antigen haplotype association with Addison's disease in three populations. Eur J Endocrinol. 2007;157(6):757-61. doi: 10.1530/EJE-07-0290.
235. Myhre AG, Undlien DE, Lovas K, Uhlving S, Nedrebo BG, Fougner KJ, et al. Autoimmune adrenocortical failure in Norway autoantibodies and human leukocyte antigen class II associations related to clinical features. J Clin Endocrinol Metab. 2002;87(2):618-23. doi: 10.1210/jcem.87.2.8192.
236. Skinningsrud B, Lie BA, Lavant E, Carlson JA, Erlich H, Akselsen HE, et al. Multiple loci in the HLA complex are associated with Addison's disease. J Clin Endocrinol Metab. 2011;96(10):E1703-8. doi: 10.1210/jc.2011-0645.
237. Muto A, Hoshino H, Madisen L, Yanai N, Obinata M, Karasuyama H, et al. Identification of Bach2 as a B-cell-specific partner for small maf proteins that negatively regulate the immunoglobulin heavy chain gene 3' enhancer. EMBO J. 1998;17(19):5734-43. doi: 10.1093/emboj/17.19.5734.
238. Muto A, Ochiai K, Kimura Y, Itoh-Nakadai A, Calame KL, Ikebe D, et al. Bach2 represses plasma cell gene regulatory network in B cells to promote antibody class switch. EMBO J. 2010;29(23):4048-61. doi: 10.1038/emboj.2010.257.
239. Roychoudhuri R, Hirahara K, Mousavi K, Clever D, Klebanoff CA, Bonelli M, et al. BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis. Nature. 2013;498(7455):506-10. doi: 10.1038/nature12199.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1328

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-319962

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2017