GENETIC RISK FACTORS IN
AUTOIMMUNE ADDISON'S DISEASE

Maribel Aranda Guillén

Stockholm 2023
All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023
© Maribel Aranda Guillén, 2023
Genetic risk factors in autoimmune Addison’s disease
Thesis for Doctoral Degree (Ph.D.)

By

Maribel Aranda Guillén

The thesis will be defended in public at Rolf Luft Auditorium, L1:00, Karolinska University Hospital, on 23 November 2023 at 09:15 AM.

Principal Supervisor:

Daniel Eriksson
Department of Medicine, Solna
Karolinska Institutet, Stockholm, Sweden
Department of Immunology, Genetics and Pathology
Uppsala University, Uppsala, Sweden

Co-supervisors:

Olle Kämpe
Department of Medicine, Solna
Karolinska Institutet, Stockholm, Sweden
Department of Endocrinology
Karolinska University Hospital, Stockholm, Sweden

Sophie Bensing
Department of Molecular Medicine and Surgery
Karolinska Institutet, Stockholm, Sweden
Department of Endocrinology
Karolinska University Hospital, Stockholm, Sweden

Eystein Sverre Husebye
Department of Clinical Science
University of Bergen, Bergen, Norway
Department of Medicine
Haukeland University Hospital, Bergen, Norway
K.G. Jebsen Center for Autoimmune Disorders
Bergen, Norway

Opponent:

Tiinamaija Tuomi
Finnish Institute for Molecular Medicine
University of Helsinki, Helsinki, Finland
Folkhälsan Research Center
Biomedicum Helsinki, Helsinki, Finland
Abdominal Centre, Endocrinology
University of Helsinki and Helsinki University Hospital, Helsinki, Finland
Lund University Diabetes Centre
Department of Clinical Sciences
Lund University, Lund, Sweden

Examination Board:

Maria Lodefalk
Department of Paediatrics
Faculty of Medicine and Health
Örebro University, Örebro, Sweden

Åsa Johansson
Department of Immunology, Genetics and Pathology
Science for Life Laboratory
Uppsala University, Uppsala, Sweden

Susanna Cardell
Department of Microbiology and Immunology
Institute of Biomedicine
University of Gothenburg, Gothenburg, Sweden
The immune system defends our organism from attacks such as viruses or bacteria, and it can decide not only to attack external agents, but also our own organism inducing an autoimmune disease.

How and why most autoimmune diseases arise is unknown but it is believed to be due to a combination of chance, genetic predisposition and environmental factors. Hence, genetic studies or research on environmental factors that contribute to the development of autoimmunity are crucial to a better understanding of its origin, thus helping to improve the diagnosis and treatment of autoimmune pathologies.

Addison's disease is an adrenal failure that is caused by autoimmunity in more than 70% of the cases, where the immune system destroys the adrenal glands located above our kidneys. These glands are responsible for the production of the steroid hormones such as aldosterone and cortisol, which regulate important functions such as our blood pressure and metabolic response to stress. An adrenal insufficiency is therefore lethal if not treated in time.

A further understanding of the pathological mechanisms that lead to autoimmune Addison’s disease can help us decipher why and how the disease arises. The analysis of the genome is a suitable approach to gain new insights, since both genetic and environmental factors are involved in the development of the disorder.

In this thesis, Paper 1 aims to investigate the genetic factors behind autoimmune Addison's disease. By combining all samples available from Sweden and Norway, we could overcome the main limitation of previous genetic studies and thus we used a method called GWAS, in which the entire genome was analyzed. Briefly, the genome of patients with Addison’s disease was compared against the genome of unaffected individuals to search for differences that may play a part in the condition. We identified nine regions in the genome associated with autoimmune Addison's disease, of which four were previously unknown. In all the identified regions, the genes are involved in the functioning of the immune system, which is consistent with the fact that immune mechanisms cause the disease.

In Paper 2, we investigated whether the nine regions, identified in Paper 1, could estimate the susceptibility of each individual to develop autoimmune Addison's disease. In other words, we calculated a risk score per each individual and evaluated whether it was a good estimation of the risk of developing the disease. We concluded that the risk score was not useful to predict whether someone in the general population will suffer from the disease. However, by closely examining the risk score in patients that were diagnosed during childhood, which is not often the case, we observed that almost all of them had a high risk score. This suggested that they carried a greater number of genetic risk factors, promoting an earlier development of the disease. In addition, two patients with early disease debut showed a low risk score for the disease. DNA analyses revealed that they had never suffered from autoimmune Addison's disease but had inherited mutations leading to other types of adrenal insufficiency not caused by autoimmunity.
The blood test to detect 21-hydroxylase autoantibodies is used to determine if an individual has autoimmune Addison's disease. The level of autoantibodies in the blood tends to become more difficult to detect the more time has passed since diagnosis, however the genetic predisposition to the disease is constant and does not diminish over the years. In Paper 3, we analyzed the risk score in patients in whom the autoantibodies were no longer detected. The risk score, or in other words, the genetic predisposition to autoimmune Addison’s disease, was low for 35 patients, which we decided to further analyze genetically. We confirmed that 5 out of the 35 also carried mutations responsible for other types of adrenal insufficiency not caused by autoimmunity. Thus, we demonstrated the potential utility of the risk score for autoimmune Addison’s disease to select patients who would benefit from genetic testing.

Establishing the etiology of Addison's disease, whether it is autoimmunity or another cause, is essential to implement a correct clinical follow-up of the disease and to provide genetic counseling if necessary. Hence, a personalized medicine approach with the risk score for autoimmune Addison’s disease as a tool, holds promise for an enhanced diagnostic accuracy of the disease.

In summary, this thesis contributes to a better understanding of the genetic factors behind the development of autoimmune Addison’s disease and its possible use in the diagnosis of the different causes of adrenal insufficiency. Future studies are needed to understand how these genetic factors promote the onset of the disease.
ABSTRACT

Autoimmune Addison’s disease is the most common form of primary adrenal insufficiency in the Western world. The low prevalence of the disease has hampered large-scale unbiased genetic studies where the entire genome could be examined at once. By combining the two largest biobanks of DNA from patients with autoimmune Addison’s disease, we identified in Paper I nine independent risk loci with a genome-wide association study of 1223 patients and 4097 geographically matched controls. These results explained up to 41% of the heritability of the disease, which, in a twin study, has been estimated to be as high as 0.97 [95% CI 0.88-0.99].

In Paper II, we derived a polygenic risk score for autoimmune Addison’s disease with the same dataset from the first study. The polygenic risk score enabled an estimation of disease susceptibility at the individual level and the discrimination of other etiologies of primary adrenal insufficiency, uncovering cases previously presumed to have the autoimmune form of Addison’s disease.

In Paper III, we explored the use of our polygenic risk score to efficiently triage patients who may benefit most from whole-genome sequencing to achieve the correct diagnosis and appropriate clinical management of the disease. Monogenic forms of primary adrenal insufficiency were found in 5 out of 35 cases with low polygenic risk score for autoimmune Addison’s disease, and we found an additional of three cases with suspected monogenic disease. This study highlights the potential of polygenic risk score as a tool to in the clinical evaluation of primary adrenal insufficiency.

In summary, this thesis sheds light on the genetic risk factors behind the development of autoimmune Addison’s disease and their potential utility as diagnostic classifiers. Future studies are warranted to further our understanding of the biological role of the associated genetic risk factors.
LIST OF SCIENTIFIC PAPERS


*# Equal contributions

GWAS for autoimmune Addison’s disease identifies multiple risk loci and highlights AIRE in disease susceptibility.


A polygenic risk score to help discriminate primary adrenal insufficiency of different etiologies.


Low polygenic risk score for autoimmune Addison’s disease identifies misdiagnosed cases of monogenic primary adrenal insufficiency.

Manuscript


LIST OF ABBREVIATIONS

AAAS  Aladin WD repeat nucleoporin
AAD  Autoimmune Addison's disease
ABCD1  ATP binding cassette subfamily D member 1
ACMG  American College of Medical Genetics and Genomics
ACTH  Adrenocorticotropic hormone
AIRE  Autoimmune regulator
APS-1  Autoimmune polyendocrine syndrome type 1
APS-2  Autoimmune polyendocrine syndrome type 2
AUC  Area under the curve
BACH2  BTB domain and CNC homolog 2
CAH  Congenital adrenal hyperplasia
CLEC16A  C-type lectin domain containing 16A
CTLA4  Cytotoxic T-lymphocyte associated protein 4
CYP11A1  Cytochrome P450 family 11 subfamily A member 1
CYP11B1  Cytochrome P450 family 11 subfamily B member 1
CYP17A1  Cytochrome P450 family 17 subfamily A member 1
CYP19A1  Cytochrome P450 family 19 subfamily A member 1
CYP21A2  Cytochrome P450 family 21 subfamily A member 2
DAX1  Dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1
ExAC  Exome Aggregation Consortium
gnomAD  Genome Aggregation Database
GPX1  Glutathione peroxidase 1
GRCh37  Genome Reference Consortium Human Build 37
GWAS  Genome-wide association study
HLA  Human leukocyte antigen
HPA  Hypothalamic-pituitary-adrenal
HSD3B2  Hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2
IGV  Integrative Genomics Viewer
LD  Linkage disequilibrium
LPP  LIM domain containing preferred translocation partner in lipoma
MC2R  Melanocortin 2 receptor
MHC  Major histocompatibility complex
MRAP  Melanocortin 2 receptor accessory protein
NNT  Nicotinamide nucleotide transhydrogenase
<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR0B1</td>
<td>Nuclear receptor subfamily 0 group B member 1</td>
</tr>
<tr>
<td>OMIM</td>
<td>Online Mendelian Inheritance in Man</td>
</tr>
<tr>
<td>PBWT</td>
<td>Positional Burrows Wheeler Transform</td>
</tr>
<tr>
<td>PEX</td>
<td>Peroxisomal biogenesis factor</td>
</tr>
<tr>
<td>PHD2</td>
<td>Plant homeodomain 2</td>
</tr>
<tr>
<td>POR</td>
<td>Cytochrome P450 oxidoreductase</td>
</tr>
<tr>
<td>PRDX3</td>
<td>Peroxiredoxin 3</td>
</tr>
<tr>
<td>PRS</td>
<td>Polygenic risk score</td>
</tr>
<tr>
<td>PTPN22</td>
<td>Protein tyrosine phosphatase non-receptor type 22</td>
</tr>
<tr>
<td>SIGLEC5</td>
<td>Sialic acid binding Ig like lectin 5</td>
</tr>
<tr>
<td>SH2B3</td>
<td>SH2B adaptor protein 3</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>STAR</td>
<td>Steroidogenic acute regulatory protein</td>
</tr>
<tr>
<td>TXNRD2</td>
<td>Thioredoxin reductase 2</td>
</tr>
<tr>
<td>UBASH3A</td>
<td>Ubiquitin associated and SH3 domain containing A</td>
</tr>
<tr>
<td>WGS</td>
<td>Whole-genome sequencing</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

Autoimmune diseases arise when the immune system targets the body’s own tissues. This is believed to stem from a complex interplay of genetic and environmental factors and stochastic events. The global incidence of many autoimmune disorders is on the rise. The exact etiological mechanisms are, however, largely unknown for the vast majority of autoimmune diseases [1].

Addison’s disease, or primary adrenal insufficiency, can be caused by monogenic disease, adrenalectomy, adrenal hemorrhage, infection or medications, but the predominant etiology present in more than 70% of cases is the autoimmune-mediated destruction of the adrenal cortex [2].

In autoimmune Addison’s disease, a heritability as high as 0.97 (95% CI 0.88-0.99) has been reported [3], highlighting the significant influence of genetic factors in its development. Nevertheless, the low prevalence of the disease, has hindered large-scale genetic studies where the entire genome could be investigated at once.

Larger sample size, in conjunction with available genetic methodology have the potential to facilitate an unbiased investigation of the genome in patients with autoimmune Addison’s disease. A deeper understanding of the genetic components associated with this disorder may provide valuable insights into the pathogenesis, and thereby targets for potential future preventative treatments. Moreover, the identification of the specific genetic risk factors enables the estimation of individual disease susceptibility through the calculation of a polygenic risk score.

A dedicated polygenic risk score might also prove instrumental for a precision medicine approach in distinguishing complex, polygenic autoimmune Addison’s disease from the rarer monogenic forms of primary adrenal insufficiency, which may be underdiagnosed among the adult population when hypomorphic genetic variants are the cause of disease.

In this thesis, we performed a genome-wide association study in autoimmune Addison’s disease that overcame the main limitations of the genetic studies performed hitherto. The identified genetic risk factors allowed us to estimate the susceptibility for the disease at the individual level by developing a polygenic risk score. This helped us to accurately diagnose patients with a monogenic cause of primary adrenal insufficiency.
2 LITERATURE REVIEW

2.1 AUTOIMMUNE DISEASES

The idea of autoimmunity or immune response directed towards the body’s self-components has a diffuse origin. In 1901, Paul Ehrlich coined the term *horror autotoxicus* to describe autoimmunity as a condition to which an organism has an innate aversion. Three years later, Julius Donath and Karl Landsteiner reported cases of autohemolysis in extremely cold conditions and described a pathological role of autoantibodies. Why this evidence was not enough to establish the concept of autoimmune disease is still being discussed. During the following years, several clinical reports and experimental data pointed towards evidence of autoimmune disorders, though this possibility was neglected until 1951, when hemolytic anemia was the first disease to be recognized as autoimmune [4, 5].

In 1957, the pathogenesis of autoimmune diseases was attributed to harmful autoantibodies and, accordingly, Witebsky established the criteria to discern autoimmune disorders [6]. Similarly to Kock’s postulates, Witebsky suggested that the presence of the disease-causing agent, in this case autoantibodies, as well as their targeted autoantigen should be observed in patients and experimental animals, and that similar pathological features must be seen in animals when inoculating the disease-causing autoantibody [6]. These criteria, however, did not consider the role of cellular immune responses in the etiology of autoimmune pathology. Hence, these postulates were reviewed by Rose and Bona, and three kind of evidence were proposed to demonstrate the autoimmune origin of a condition: 1) a direct evidence by inducing the disease with the transfer of pathogenic autoantibodies or T-cells on experimental animals, by placental transfer, or even human volunteers, 2) an indirect evidence by isolating harmful autoantibodies or T-cells from patients or by inducing disease on experimental animals with autoantigens and, 3) circumstantial evidence based on autoimmune family history, lymphocytic infiltration in the affected organ, association with the major histocompatibility complex (MHC) and positive response to immunosuppressive therapy [7].

Autoantibodies are important in some but not all autoimmune diseases. The passive placental transfer of autoantibodies during pregnancy is a natural experiment that demonstrates this. In diseases caused by pathogenic autoantibodies, like myasthenia gravis and Graves’ disease, the newborn may be severely affected until the maternal autoantibodies have been cleared [8, 9].

Most of autoimmune diseases are, however, driven by autoreactive T-cells, in which the role of autoantibodies is not completely understood. In diseases like type I diabetes and autoimmune Addison’s disease, mothers that suffer from T-cell driven autoimmune disorders give birth to unaffected children [7].

The etiology of most autoimmune disorders remains a mystery. They are believed to result from a complex interplay of genetics and environmental factors and stochastic events. When autoreactive T- and/or B-cells escape the checkpoints of both central and peripheral immune tolerance, there is a break of tolerance that leads to autoimmunity. Then, an autoimmune disorder is developed by
genetically susceptible individuals that encounter certain environmental factors or stochastic events [10-12].

As a whole, autoimmune diseases are estimated to affect 3-5% of the general population [12-15]. In other words, between 234-390 million people worldwide suffer from a chronic autoimmune condition and their incidence is increasing [16]. More often than expected by chance, autoimmune diseases aggregate in families and an individual patient often suffers from more than one autoimmune disorder, indicating shared risk factors between different autoimmune diseases [17, 18]. As evident from heritability studies, the relative importance of genetic and environmental risk factors varies among diseases and populations. Moreover, depending on the geographical area, the prevalence of individual autoimmune disorders varies indicating that the underlying risk factors vary in frequency between different countries [12-15]. Studies on migrating populations have been conducted to explore whether the prevalence of autoimmune diseases is influenced by a new environment. For example, a study on type I diabetes in families migrating from Sardinia or continental Italy to Germany, showed that the prevalence of disease in children remained unchanged, reinforcing the role of genetic factors in disease development [19, 20]. Conversely, an increase in prevalence of multiple sclerosis was observed in children migrating from low to high-risk countries, suggesting the importance of the environment [21].

Genetic risk factors are of great importance for the development of autoimmune disease and the study of monogenic forms of autoimmune disorders has unveiled fundamental physiological and pathological immunological pathways [22]. Nevertheless, for most autoimmune conditions, several genes are involved in their development. For instance, the most common and often strongest genetic risk factor in autoimmune diseases is the MHC-region on chromosome 6, but additional risk loci are required for the development of disease. Risk alleles in autoimmune disorders may be associated with several autoimmune conditions, which might point out shared pathological mechanisms or a potential pleiotropy of causal variants, that is, the same variant affects more than one condition [12, 23]. For instance, the same genetic variant in \textit{PTPN22} gene has been found to increase the susceptibility to a number of different complex autoimmune diseases like, among others, type I diabetes, Hashimoto's thyroiditis and vitiligo [24-26].

The predominant approach for studying the genetic architecture of autoimmune diseases is through case-control association studies, which suffer from the same limitations as any other non-genetic epidemiological study. For example, the introduction of bias by a factor, often referred to as a confounder, which influences both the genes under examination and the disease of interest without serving as an intermediate link between them, can result in what is known as confounding, which should be controlled for during the genetic analyses [27].
2.2 MENDELIAN VS COMPLEX GENETIC DISEASES

Monogenic diseases are caused by mutations in a single gene that segregates in families following a Mendelian inheritance pattern. Historically, linkage analysis, or genetic mapping by studying the cosegregation of causal variants within pedigrees, were successful to identify inborn errors responsible for monogenic disorders [28-31]. Today, sequencing studies of selected cases with severe phenotype or early onset has identified the etiology of many monogenic traits, typically caused by a high-impact mutation [32, 33]. Sequencing of a trio consisting of an affected patient and two unaffected parents, may be enough to pinpoint an underlying mutation [34]. An example of autoimmune monogenic disease is the autoimmune polyendocrine syndrome type-1 (APS-1). This is an ultra-rare recessive disorder caused uniquely by mutations in the AIRE gene on chromosome 21 involved in the thymic negative selection during T-cell development. Patients with autoimmune manifestations due to an AIRE gene mutation with dominant inheritance has also been reported [35, 36].

In contrast to monogenic disease, in complex genetic diseases no high-impact mutation is by itself enough to cause the disease, and no single genetic risk factor is strictly required. Hence, complex diseases do not follow a simple Mendelian inheritance pattern, which makes difficult to determine the risk of inheriting or passing on these disorders [23, 28, 31]. Examples of complex autoimmune disease are type 1 diabetes, vitiligo or pernicious anemia. In complex disease, sequencing of a small number of subjects will not be successful in identifying the genetic causes, due to polygenicity, genetic heterogeneity and incomplete penetrance. Common variants with low effect size are known to contribute to the genetic architecture of complex diseases and thousands of genome loci have been successfully associated with hundreds of traits and diseases by means of case-control association studies [23, 29].

2.2.1 Candidate-gene studies

In the last almost 20 years, genome-wide association studies have been the method of choice to study common genetic risk factors in common complex diseases [27, 37, 38]. However, when studying rare diseases, the small sample size has limited the scope of the studies to candidate genes – typically one gene at a time. Typically, the candidate-gene approach compares the frequency of genetic variants of a single gene between cases and controls. Unfortunately, this strategy offers no control over population stratification, which refers to different proportions of ancestries in affected and unaffected groups in association studies, and that may have a confounding effect. In other words, evolutionary forces such as mutation, natural selection, recombination and genetic drift, have differently shaped the allele frequencies of genetic variants among geographical regions [39]. Hence, genetic variants more common in certain areas may be spuriously associated with disease if different ancestries differ between cases and controls. Taken together, a small sample size and analyses restrained to only parts of the genome limits the chances to detect new risk loci, while population substructure increases the risk for false positive associations [40].
2.2.2 Genome-wide association studies

A genome-wide association study (GWAS) is a case-control experimental design that also investigates the difference in frequency of genetic variants among affected individuals versus unaffected [23, 29, 41]. In order to associate a genetic variant to a trait, a significant statistical difference in frequency between these groups must exist. GWAS relies on the tendency of DNA variants to be inherited together in linkage disequilibrium (LD) blocks, which enables tagging of genetic variability across the whole genome with a limited number of genetic markers [23, 29, 42]. Thereby, common disease-causing genetic variants are expected to be tagged by genotyped variants. A common misconception, however, is the equalization between association and causality. Associated variants are simply close enough to the causal variant to be inherited together and produce a significant result. Therefore, genetic variants associated with a disease with a GWAS are not necessarily, and even unlikely, the disease-causing variants [23, 29, 42].

Typically, the genetic markers selected to be on a genotyping chip are common in the general population [23, 29, 43]. Uncommon variants tend to have lower LD correlations with the genotyped markers due to allele frequency disparity, reducing the power to associate them with disease [29]. Theoretically, larger sample sizes could overcome the low power to associate rare variants with complex diseases [44].

Single nucleotide polymorphism (SNP) arrays are used to genotype selected genetic markers, though the lower cost of whole-genome sequencing (WGS) provides an increasing availability of this kind of data to use in GWAS [23]. WGS provides nearly every single nucleotide in a genome, increasing considerably the density of markers and avoiding dependency on LD blocks and tagging SNPs. While high throughput sequencing for GWAS gets more widely spread, imputation is currently used to increase the density of markers. Imputation consists of predicting non-genotyped markers by inferring the haplotypes they belong to and extracting the information from haplotypes observed in a fully sequenced reference panel [23, 45].

GWAS has explained a proportion of heritability in many complex traits. However, unexplained heritability still remains and may be further investigated with a combination of sequencing data and larger sample sizes [23, 43, 46]. Rare and copy number variants with moderate to high effect size have often been overlooked in previous association studies [43].

2.2.2.1 Polygenic risk score

Besides facilitating a better pathobiological understanding, GWASs are thought to provide a potent base for preventive and personalized medicine by predicting the risk for disease development [23, 47]. A polygenic risk score (PRS), representing the genetic burden inherited by an individual, can be calculated with summary statistics from GWAS and individual-level genotype data. A PRS is the sum of the number of risk alleles that an individual carries weighted by their corresponding estimated effect size [47].

The simplicity of a polygenic risk score can be a double-edged sword. On one hand, PRS allows to stratification of individuals into low and high risk from a population perspective. On the other hand,
the potential of risk stratification may be misconceived when applying the uncertain risk estimates to single individuals in clinical applications [48, 49]. For the vast majority of diseases, a PRS does not explain the entire genetic component and does not consider environmental factors and potential environment-gene or gene-gene interactions.

PRS has been commercialized in private testing companies to screen embryos prior to in vitro fertilization even if no clinical studies for such aim have been undertaken [50]. However, by focusing on individuals with the highest or lowest risk for a trait, PRS has potential in preventive strategies or diagnostic workup as a complement to already existing clinical guidelines [44, 51]. For instance, a PRS for breast cancer have been suggested as a potential tool to refine screening programs by identifying patients at risk, while reducing overdiagnosis in women with low genetic predisposition to the disease [52, 53]. Also, low PRS for common complex diseases have been postulated as a possible indicator to triage patients that are most to take profit from sequencing for monogenic disease diagnosis [54].

2.2.3 Next generation sequencing

Since the first human genome was sequenced in 2001 [55, 56], great technical improvements and decreasing sequencing costs have allowed researchers and physicians to access high-throughput sequencing [32]. This method consists of the fragmentation of DNA molecules in short segments that are sequenced simultaneously. Sequencing reads are aligned to a reference genome and any mismatches are listed as genetic variants.

Next generation sequencing of the entire genome is WGS. Besides providing nearly every single nucleotide and, therefore, more dense data as compared to GWAS genotyping, WGS allows for the investigation of structural variants, such as deletions or insertions, and unique variants that are often not detected by SNP arrays [33, 43].

WGS is the method of choice for identifying high-impact causal variants in monogenic diseases [32, 57]. It remains challenging, however, to identify disease-causing variants in the myriad of variation that makes up a human genome [58]. In an attempt to standardize the interpretation of genetic variants, a workgroup including the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology published a guideline in 2015 to simplify and systematize the classification of genetic variants into five categories: pathogenic, likely pathogenic, uncertain significance, likely benign and benign [59]. The criteria were based on different levels of evidence such the effect of a variant (e.g. destruction of the gene), de novo variants, allele frequency in the general population, and segregation data [60].

Common variants in the general population are unlikely to be responsible for Mendelian diseases. ACMG classification criteria relies heavily on publicly available population databases where extensive sequencing efforts have listed observed human variation and their frequencies across different ancestries. ExAC (Exome Aggregation Consortium) was the first database large enough to evaluate recurrence frequency of rare variants that previously were assumed to cause monogenic diseases because of their apparent absence in the population [61]. Indeed, among the more than 60,000
exomes that ExAC aggregated, 54 variants previously classified as pathogenic were found on average in each individual, demonstrating their high frequency and suggesting their misclassification [62]. For example, Walsh et al. demonstrated that rare variants thought to be causal for cardiomyopathy were actually too common in the general population to be the main and only cause of a Mendelian disease [63].

2.2.4 Heritability

The term ‘heritability’ is often used interchangeably to refer to two very different concepts: (1) the quality of being inheritable or congenital, and (2) the population-specific parameter between 0 and 1 that reflects how much a phenotype is attributable to the genotype. Although complex diseases are not inheritable in the Mendelian sense of the word, heritability as a population parameter is a useful measure to help identify complex diseases where genetic studies could be productive. The population parameter (2) is calculated by dividing the genetic variance of a trait by the total variance [64].

2.3 ADDISON’S DISEASE

The adrenal cortex is responsible for the production and secretion of mineralocorticoids such as aldosterone, glucocorticoids, of which cortisol is the most important, and androgens. The release of cortisol is regulated by the hypothalamic-pituitary-adrenal (HPA) axis [65]. In physiological conditions, the hypothalamus secretes corticotrophin-releasing hormone that stimulates the pituitary to synthesize adrenocorticotropic hormone (ACTH), responsible for activating the adrenal cortex to produce glucocorticoids [66, 67]. Defects at any level of the HPA axis leads to adrenal insufficiency, categorized as primary, secondary or tertiary depending on whether the failure is on the level of the adrenal, the pituitary or the hypothalamus, respectively [65].

Addison’s disease or primary adrenal insufficiency is characterized by the malfunction of the adrenal cortex. This condition often leads to an insufficient production of glucocorticoids, mineralocorticoids and adrenal androgens. If replacement therapy is not supplied this may be lethal [2, 68-72]. Impaired quality of life and risk of premature death is reported in spite of an effective treatment [73, 74]. It is a rare disorder, the highest prevalence rate has been estimated in Iceland with 221 cases per million in the adult population and 140 cases per million in Norway, while 5 and 4 cases per million have been reported in Japan and Korea, respectively [75-78].

In line with what Thomas Addison first described in 1855, the typical symptomatology is fatigue, weakness, salt craving, abdominal pain, vomiting and anorexia [2, 68, 71, 72, 79, 80]. This non-specific symptomatology often contributes to a delayed diagnosis that increases the risk of an adrenal crisis [69, 74]. It has been reported that about half of patients received their diagnose right
after an adrenal crisis [81], which is characterized by acute disturbances of hemodynamics including severe hypotension, nausea, vomiting, hyponatremia and eventually hypovolemic shock [82]. Elevated levels of adrenocorticotrophic hormone (ACTH) accompanied by low serum levels of cortisol confirm the diagnosis of Addison’s disease, but in case of inconclusive results a Synachten test is performed [68, 83, 84]. Clinical signs at diagnosis may include hyponatremia, hyperkalemia and skin hyperpigmentation [66, 69].

Addison’s disease is often used as synonym of autoimmune Addison’s disease (AAD), since the main cause of a primary adrenal insufficiency in adults is autoimmunity, accounting for more than 70% of cases [85-87]. Nevertheless, until the middle of the 20th century, tuberculosis used to be the most common cause in the Western world [78, 88-90]. Other minor causes are adrenalectomy, medication, infection, adrenal hemorrhage, infiltration and monogenic forms (Table 1) [70, 73, 74, 84, 90]. Clinical guidelines highlight the relevance of investigating Addison’s disease etiology once the diagnosis is confirmed to promptly provide an adequate management and, if appropriate, genetic counselling [2, 66, 68, 83, 91, 92].

2.3.1 Autoimmune Addison’s disease

In AAD, a cellular immune response is responsible for the destruction of the adrenal cortex [93]. The steroidogenic enzyme 21-hydroxylase is a highly specific tissue-restricted autoantigen and the presence of autoantibodies against 21-hydroxylase is used for AAD diagnosis [94, 95]. Autoantibodies against 21-hydroxylase can be found in patients with overt AAD, but also during the preclinical stage of the disease [71, 96]. The typical onset of AAD is at the age of 20-50 years, though a few cases develop the disease during childhood [2, 73].

Between 60-80% of AAD patients present other autoimmune comorbidities. In fact, AAD is a component of autoimmune polyendocrine syndromes type 1 and 2 (APS-1 and APS-2, respectively) [71, 73, 90, 97]. More common than not, AAD is manifested as a part of APS-2, a syndrome of autoimmune diseases with complex inheritance. APS-2 is defined by the combination of AAD with at least one of the more common endocrine autoimmune diseases, like type I diabetes or autoimmune thyroid disease [35, 73, 90].

The heritability of AAD, either occurring in isolation or as part of APS-2, has been reported in a disease concordance study in Swedish twins to be as high as 0.97 (95% CI 0.88-0.99) [3]. It is therefore clear that genetic risk factors are involved in disease development. As a complex disease, AAD presents genetic heterogeneity, tends to appear aggregated in families [86, 87], and does not follow a simple Mendelian inheritance pattern [3, 98, 99].

Several genetic factors have been associated with AAD development. Due to the rarity of the disorder, mostly candidate gene studies have been feasible hitherto, in which a strong association with the human leukocyte antigen (HLA) complex have been established [97, 100-102]. Other non-HLA loci have also been suggested, most of them implicated in immunological pathways and associated with other autoimmune diseases like type 1 diabetes [90, 97, 98, 103, 104]. However, more recent studies have not found support for several of previously reported candidate gene
findings, probably due to intrinsic limitations of the candidate gene-approach. In most studies, a restricted number of nucleotide polymorphisms were interrogated in a limited number of samples.

In 2016, a large-scale targeted sequencing study analyzed 1853 genes in 479 AAD cases and 1394 controls, trying to capture associated genetic variants within the investigated genes [99]. By capturing the exons as well as promoters, regulatory elements and splice sites, the strong effect of HLA in AAD development was validated and a novel locus was identified in the BACH2 gene, which is implicated in B-cell functionality and T-cell development [99]. In a follow up study, the addition of another 1000 healthy control samples increased the statistical power and led to the association of common variants in AIRE gene with AAD in non-APS1 cases [105]. Still, only a small percentage of the human genome was explored in these studies, and the low sample size made the studies underpowered.

2.3.2 Monogenic forms of Addison’s disease

Addison’s disease in children is often caused by monogenic defects [66, 90, 104]. Depending on the pathological mechanism caused by genetic alterations, monogenic forms of Addison’s disease can be classified in six main groups: (1) defects of steroidogenesis, due to mutations in any of the genes that encode for steroidogenic enzymes and their cofactors, leading to their deficiency and consequently to congenital adrenal hyperplasia; (2) adrenal dysgenesis, owing to mutations in crucial genes for adrenal gland development that leads to adrenal hypoplasia; (3) ACTH resistance / familial glucocorticoid deficiency, when genetic variants affect the ACTH signaling; (4) impaired adrenal redox homeostasis; (5) metabolic disorders leading to adrenal lesions and (6) autoimmune [90, 91, 106, 107]. To date, APS-1, caused by mutations in AIRE is the only monogenic autoimmunity syndrome known to feature Addison’s disease as a component.

2.3.2.1 Congenital adrenal hyperplasia

Defects in enzymes involved in the production of glucocorticoids and/or mineralocorticoids form a group of disorders recessively inherited collectively called congenital adrenal hyperplasia (CAH, Figure 1). Over 95% of cases are attributed to genetic defects in the gene CYP21A2 encoding 21-hydroxylase, relevant for both glucocorticoid and mineralocorticoid production. Less common genetic defects occur in other genes involved in steroidogenesis, such as STAR, CYP11A1, HSD3B2, POR, CYP17A1 or CYP11B1 [108-110].

The initial steps of cholesterol conversion into steroid hormones are catalyzed proteins encoded by STAR and CYP11A1. Hence, a deficiency in any of these enzymes results in an insufficiency of all kinds of steroid hormones and manifests similarly biochemically and clinically [109, 111]. STAR gene encodes for the steroidogenic acute regulatory protein, responsible for the transport of cholesterol from the outer to the inner mitochondrial membrane. CYP11A1 encodes the side-chain cleavage enzyme that catalyzes the conversion of cholesterol to pregnenolone, the rate-limiting step of steroidogenesis. A lack of StAR leads to an accumulation of lipid droplets and adrenal enlargement, why StAR deficiency is known as lipoid congenital adrenal hyperplasia [109-111]. In contrast, adrenal atrophy is observed when CYP11A1 is defective. Both genes have been associated with classic and
non-classic forms of CAH, being the non-classic characterized by a later age of onset and hypomorphic genetic variants that enable the retention of up to 38% enzyme activity [91, 109, 111].

Besides disease-causing variants in steroidogenic enzymes, genetic defects in their cofactor POR have been also described. POR gene encodes for the cytochrome P450 oxidoreductase, which is a redox partner for several steroidogenic enzymes encoded by CYP17A1, CYP21A2 and CYP19A1 that depend on the electron transfer from POR to properly function. Glucocorticoid pathway is typically more affected, leading to hypertension in most of the patients [106, 109].

Figure 1. Adrenal steroidogenesis. The stepwise conversion of cholesterol is finely regulated by the steroidogenic enzymes (blue) to produce mineralocorticoids such as aldosterone in the zona glomerulosa, glucocorticoids such as cortisol in the zona fasciculata, and androgens in the zona reticularis. The enzymes require cofactors (orange) to function properly. Adapted from [67].

2.3.2.2 Adrenal dysgenesis

Among genes whose defect lead to adrenal hypoplasia congenita (Table 1), the most common is found in NR0B1, which encodes for the dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1 (DAX-1) protein [106, 107]. DAX-1 is an orphan nuclear receptor involved in adrenogonadal development. Pathogenic variants in these gene gives a distinct phenotype of adrenal dysfunction combined with hypogonadotropic hypogonadism [91, 106, 107, 110, 112].

2.3.2.3 ACTH resistance

As part of the HPA axis, ACTH signaling in the adrenal cortex stimulates both the proliferation of adrenocortical cells and the production of glucocorticoids. Mineralocorticoids are instead regulated by the renin-angiotensin-aldosterone loop [107]. Hence, defects in ACTH signaling tend to affect
solely to glucocorticoid production and therefore ACTH resistance and familial glucocorticoid deficiency are terms used interchangeably. Pathological genetic variants in genes coding for ACTH receptor (MC2R) and the accessory protein MRAP cause primary adrenal insufficiency by unresponsiveness to ACTH [106, 107, 110].

2.3.2.4 Impaired adrenal redox homeostasis

Adrenal steroidogenesis requires intense mitochondrial metabolic activity that generates reactive oxygen species. Increased levels of these species inhibit steroid hormone production, and therefore, a balanced redox potential is key to prevent an early stop of steroidogenesis [106]. Genetic variants associated with primary adrenal insufficiency have also been described in the genes NNT, TXNRD2, AAAS, GPX1 and PRDX3, all of them involved in regulating the levels of reactive oxygen species in the cell [106, 107, 110, 113].

2.3.2.5 Metabolic disorders

Beyond disorders related to cholesterol synthesis and lysosomal and mitochondrial defects, one of the prevalent metabolic disorders that lead to primary adrenal insufficiency is due to pathological variants in the ABCD1 gene, which encodes for a transporter located in the membrane of peroxisomes. A dysfunction of the transporter hinders very-long-chain fatty acids from being degraded in the peroxisome and subsequently they accumulate, among other organs, in the adrenal glands [90]. The clinical condition is known as adrenoleukodystrophy (OMIM# 300100), characterized by progressive neurodegeneration and adrenal failure. Strikingly, the clinical manifestations are extremely heterogeneous and a correlation between genotype and phenotype is inexistent [110, 114, 115]. Less common defects in the peroxisomal metabolism associated with primary adrenal insufficiency are due to genetic variants in the PEX gene family [2, 106].

2.3.2.6 Autoimmune syndrome

APS-1 is a monogenic disorder caused by a defective AIRE gene. During T-lymphocyte development in medullary thymic cells, the AIRE protein is involved in negative selection of autoreactive T-cells by promoting the ectopic expression of self-antigens. T-cells that interact strongly with presented self-proteins die by apoptosis or become regulatory T-cells. Thus, the loss of AIRE function leads to an escape of autoreactive T-cells into circulation, and an altered regulatory T-cell repertoire. The resulting clinical picture includes tissue-specific autoimmune disorders [35, 90]. A clinical diagnosis of APS-1 is given when at least two of the following conditions are manifested: AAD, hypoparathyroidism and chronic mucocutaneous candidiasis. Other associated manifestations such as vitiligo, primary ovarian failure, alopecia areata, type I diabetes, autoimmune thyroid disease, enamel hypoplasia, nail dystrophy and a skin rash might be observed. In addition, autoantibodies against interferon-α, interferon-ω and interleukin-22 are found in almost all patients [35, 90, 116]. The prevalence of APS-1 is estimated to be 1 case per 100,000 individuals [35].
2.3.3 Other causes of Addison’s disease

Besides autoimmunity and monogenic disease, adrenalectomy, adrenal hemorrhage, infection, infiltration, and certain medications can cause primary adrenal insufficiency [2, 84, 104].

Examples of medications associated with Addison’s disease are ketoconazole, fluconazole, etomidate, metyrapone, itraconazole, trilostane that can inhibit steroidogenic enzymes, whereas mitotane, rifampicin, troglitazone, phenytoin and phenobarbital are known to induce P450-cytochrome enzymes that increase cortisol metabolism. Anticoagulants, for example heparin, can induce adrenal hemorrhage [2, 65, 84, 90, 117].

Checkpoint inhibitors have revolutionized the treatment of certain malignancies when removing the natural breaks of immune response to treat tumors more effectively. The other side of the coin, though, is the emergence of autoimmune adverse events. A few case-reports have linked the use of checkpoint inhibitors to primary adrenal insufficiency [74, 118-120]. In an attempt to review primary adrenal insufficiency cases due to checkpoint inhibitors, Grouthier et al. screened the World Health Organization’s pharmacovigilance database and found 45 confirmed cases of primary adrenal insufficiency out of 50,108 subjects that experienced adverse events after therapy as of 2019 [118].
Table 1. Summary of causes of primary adrenal insufficiency. The predominant etiology of Addison’s disease is autoimmunity, though other rarer causes are monogenic disease, haemorrhage, adrenalectomy, infection, infiltration and medicines. Adapted from [2, 70, 72, 84, 106].

<table>
<thead>
<tr>
<th>Etiology</th>
<th>Group of monogenic disease</th>
<th>Disorder</th>
<th>Gene (OMIM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autoimmune</td>
<td>Autoimmune Addison’s disease</td>
<td>Autoimmune Addison’s disease</td>
<td>Associations with HLA, BACH2, AIRE, PTPN22, CTLA4, CLEC16A</td>
</tr>
<tr>
<td>Monogenic</td>
<td>Congenital adrenal hyperplasia</td>
<td>Lipoid congenital adrenal hyperplasia, P450 side chain cleavage enzyme deficiency</td>
<td>STAR (201710)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3β-hydroxysteroid dehydrogenase deficiency</td>
<td>CYP11A1 (613743)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>21-hydroxylase deficiency</td>
<td>CYP21A1 (201910)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11β-Hydroxylase deficiency</td>
<td>CYP11B1 (206010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17α-hydroxylase deficiency</td>
<td>CYP11B1 (202110)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11β-hydroxylase deficiency, 21-hydroxylase deficiency</td>
<td>CYP11B1 (202110), CYP21A2 (201910)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hydroxylases deficiency</td>
<td>CYP21A2 (201910)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aldosterone synthase deficiency</td>
<td>CYP11B2 (203400), PRKCB (607398)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Corticosterone reductase deficiency</td>
<td>CYP11B2 (203400), PRKCB (607398)</td>
</tr>
<tr>
<td>Adrenal dysgenesis</td>
<td>X-linked adrenal hypoplasia congenita</td>
<td></td>
<td>MEOX1 (617050)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MIRAGE syndrome</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Meckel syndrome</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pena-Shokeir syndrome</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pseudotrisomy 13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hydroxalethalus syndrome</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Galloway-Mowat syndrome</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ACTH resistance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Familial glucocorticoid deficiency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Impaired adrenal redox homeostasis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Metabolic disorders</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Autoimmune Addison’s disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Autoimmune Addison’s disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adrenal haemorrhage, antiphospholipid syndrome, anticogulants</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Infantile Refsum disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sitosterolemia</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kearns-Sayre syndrome</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spilins-1-phosphatase deficiency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sphingosine-1-phosphate lyase 1 deficiency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sphingosine-1-phosphatase deficiency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LXR2 (202200), MRAP (607398)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adrenal haemorrhage, antiphospholipid syndrome, anticoagulants</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Septic shock, Waterhouse-Friderichsen syndrome, tuberculosis, fungal infections, syphilis, virus such as cytomegalovirus or HIV-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tumor, metastasis, primary adrenal lymphoma, amyloidosis, sarcoidosis, hemochromatosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adrenostatic agents (e.g. ketoconazole, fluconazole, etomidate, metyrapone), mitotane, drugs that increase cortisol metabolism (e.g. rifampicin, troglitazone, phenytoin, phenobarbital), checkpoint inhibitors</td>
</tr>
</tbody>
</table>

Table 1. Summary of causes of primary adrenal insufficiency. The predominant etiology of Addison’s disease is autoimmunity, though other rarer causes are monogenic disease, haemorrhage, adrenalectomy, infection, infiltration and medicines. Adapted from [2, 70, 72, 84, 106].
3 RESEARCH AIMS

The main aim of this thesis was to investigate the genetic background of autoimmune Addison’s disease. The high heritability indicated the possibility for genetic studies to offer insights into the pathogenesis of the autoimmune destruction of the adrenal cortex.

The specific aims of the thesis were:

- To explore the common genetic variation underlying AAD, across the whole genome and with unprecedented statistical power (Paper I).
- To investigate the utility of a polygenic risk score for AAD to estimate genetic susceptibility at the individual level and discriminate pediatric patients with other disease etiologies (Paper II).
- To assess the utility of a polygenic risk score for AAD to help identify patients that may benefit from whole genome sequencing to establish a correct molecular diagnose (Paper III).
4 MATERIALS AND METHODS

Overview of the critical methodology used in this thesis. Please, refer to the constituent articles of the thesis for detailed descriptions.

4.1 SUBJECTS

Samples from patients with primary adrenal insufficiency were retrieved from the Swedish Addison Registry and the Norwegian registry for organ-specific autoimmune diseases. Both registries are national multicenter studies containing clinical, serological, and genetic data as well as serum and DNA samples from patients diagnosed with Addison’s disease [86, 87]. Healthy controls were sampled across Sweden and Norway from blood donor centers. Informed consent was obtained from all research subjects.

4.2 GENOME-WIDE ASSOCIATION STUDY

A total of 6112 samples were genotyped simultaneously with Illumina Infinium Global Screening Array 1.0 covering 692 367 genetic markers. In short, low quality samples and markers were excluded based on, among other parameters, missing data, heterozygosity, Hardy-Weinberg equilibrium, and discordant genotypes. High quality samples and markers were kept after quality control using PLINK version 1.9 [121].

Subsequently, we performed haplotype estimation (phasing) and SNP imputation to increase the breadth of the data for association analysis. The Haplotype Reference Consortium reference panel [122] and pedigree information were used to estimate haplotypes in the software SHAPEIT [123], which then were used to impute unobserved genotypes with the Sanger Imputation Service (PBWT). Genotyped and imputed markers with minor allele frequency higher than 1% were included in the association study.

Analyses were limited to unrelated European samples to avoid confounding bias due to population substructure. Finally, only cases with positive autoantibody test against 21-hydroxylase were included in the analysis to ensure a homogeneous phenotype, amounting to a total of 1223 cases and 4097 geographically matched controls. Logistic regression on allele dosages with sex and the first five principal components as covariates was performed to obtain association statistics. The HLA region was analyzed separately in an attempt to finely dissect the association of this highly variable region with autoimmune Addison’s disease. In short, the HLA region was imputed using HIBAG [124] and SNP2HLA [125] and a step-wise regression with the same covariates as the main analysis was used to identify the alleles and amino acids at HLA class I and II that best formed a model of association with the disease.
4.3 POLYGENIC RISK SCORE

By means of the individual-level genotype data from GWAS, we constructed a polygenic risk score for autoimmune Addison’s disease with the classical method of clumping and thresholding. In other words, the most significantly and independently associated SNPs in sliding windows of 250 kilobase pairs were selected for inclusion in the PRS calculation (clumping), and different levels of significance were tested, varying the number of significant SNPs included in PRS (thresholding). To avoid overfitting, the GWAS dataset was divided into a training, tuning and validation set. Association statistics from a GWAS performed within only the training set were obtained using the same methodology as in the main GWAS of Paper I. The obtained summary statistics were subsequently used to calculate PRS with genotype data on the tuning set, where thresholding was applied to refine the PRS. The most optimal configuration to calculate PRS for autoimmune Addison’s disease (PRS14$_{AAD}$) was employed on the independent validation set to evaluate its performance. PRS14$_{AAD}$ contained 5 SNPs from non-HLA regions determined with the software PRSice-2 [126, 127] and 9 HLA amino acids and alleles obtained from stepwise regression on the training set.

4.4 WHOLE-GENOME SEQUENCING, DATA ANALYSIS AND INTERPRETATION

DNA from whole blood was extracted and sent for whole-genome sequencing to the National Genomics Infrastructure and Clinical Genomics platform at the Science for Life Laboratory, Stockholm, Sweden. Pair-end 2x150 base pair sequencing was performed with TruSeq PCR-free library in Illumina HiSeq X and NovaSeq 6000. Mutation Identification Pipeline [57, 128] was used to process the sequencing data. In short, the pipeline included quality control of raw sequences, aligning of sequencing reads to reference genome (GRCh37), variant calling and recalibration, and annotation of genetic variants with public databases such as Variant Effector Predictor. Subsequently, variants were prioritized according to their deleteriousness, allele frequency, variant quality, gene intolerance prediction, inheritance mode and conservation rate. Ranked variants were then uploaded to the graphical user interface Scout [129] for further assessment. Integrative Genomics Viewer (IGV) [130] was employed to rule out potential sequencing artefacts. We examined variants in genes previously associated with primary adrenal insufficiency or present in exonic and splicing regions of any gene with an allele frequency lower than 2% in the public Genome Aggregation Database (gnomAD, version 2.1.1) [131]. We assessed variant pathogenicity with ACMG guidelines [59].

4.5 ETHICAL CONSIDERATIONS

The studies included in this thesis were approved by the ethics committees in Sweden (dnr 2008/296-31/2) and Norway (biobank 2013-1504, project 2017-624) and performed in accordance with the Declaration of Helsinki. Clinical and genetic information was pseudonymized and kept in secure infrastructures accessed only by authorized personal in order to respect the integrity and the right to privacy of the research subjects. Collected samples consisted of whole peripheral blood,
which might produce pain and hematomas in research subjects. To minimize these risks, samples were obtained during routine clinical practice. In the published studies, individual genotyping data was not provided to the journal where the manuscripts were submitted to preserve the integrity of research participants. For all studies, patients provided a written informed consent that included the risks and rights in their participation [132]. Additionally, for whole-genome sequencing analysis, patients are duly informed that if incidental findings are found, genetic counselling will be provided as ACMG recommends [133]. In instances where variants of unknown significance are found, the ethical permits allowed us to contact patients, aiming to better elucidate the relevance of the variants in the studied phenotype. Since these uncertain findings could cause unnecessary concerns, a thorough analysis of these variants was performed to make an informed decision whether contact patients or not [133]. Finally, an ethical amendment was requested to be able to report results of the studies at the individual rather than at the group level, risking compromising the anonymity of single individuals. However, the data was presented publicly in a way that it was not possible to identify the research subjects and only a small group of authorized personal had access to single individual data.
5 RESULTS AND DISCUSSION

5.1 PAPER I

By combining the world’s two largest collections of DNA from patients with Addison’s disease, we could perform a GWAS including 1223 21-hydroxylase autoantibody positive patients and 4097 geographically matched controls.

After genotyping, almost 700,000 common markers were available for association testing, that increased to more than 7 million after phasing and imputation with the Haplotype Reference Consortium. Logistic regression on allele dosages was performed with sex and the first five principal components as covariates.

Nine independent risk loci were genome-wide significantly associated with the development of AAD (Figure 2). The strongest risk loci was the HLA region, which had been previously reported [99-102], but variants in or close to PTPN22, CTLA4, BACH2, AIRE, LPP, SH2B3, SIGLEC5, and UBASH3A were also significant, of which the last four were novel in AAD [99, 105, 134-137]. Regarding other loci previously associated with AAD, only the one around CLEC16A gene showed a sub-significant genome-wide p-value, which maybe could have been significant with a larger sample size.

Figure 2. Manhattan plot showing the genetic risk factors associated with AAD in the GWAS with 1223 seropositive cases and 4097 geographically matched controls. All SNPs are represented in the figure across all chromosomes and those associated with AAD are above the dotted red line representing the genome-wide significance level. The y-axis was modified to show the level of association for the top SNP in the HLA region. Figure from Paper I [138].

Long-range linkage disequilibrium and allele heterogeneity makes difficult a proper dissection of significant associations within the HLA region. With stepwise logistic regression, seven HLA alleles and amino acids were associated with AAD, being HLA-DQB1*02:01 and HLA-DQB1*03:02 the strongest signals as previously documented [99, 139].
Strikingly, the novel and lead SNP in the AIRE locus was a missense variant (c.1411C>T, p.R471C) that replace an arginine by a cysteine surrounding a zinc ion in the PHD2 domain of the protein, which already has four cysteines in close vicinity. It is easy to speculate that this subtle change might alter the protein structure and function, though not enough to cause disease by itself. By conditional regression, an independent additional signal was identified in the AIRE gene. The association of both significant signals was best explained by an additive model, instead of recessive as in APS-1. At the time this study was published, the lead signal in AIRE seemed to be highly specific only for AAD, implicating the break of central tolerance mechanisms in the pathogenesis of the disease. Intriguingly, the same coding variant has been associated with other organ-specific autoimmune disease such as type I diabetes [140] and pernicious anemia [141] after our publication, reflecting pleiotropy.

In Paper I we could explain a great portion of AAD heritability in Scandinavia, up to 41%. This can be in part because AAD has a strong genetic component comprised of common genetic variation, and part due to an exceptionally precise phenotyping allowed by specific serological and biochemical characteristics of the patients, which increased the homogeneity of the research subjects, and carefully matched geographically controls, which helped to reduce the risk of false positive associations [39]. Nevertheless, there is still a great portion of missing heritability that might be due to other common variants with low effect size yet to be discovered by increasing sample size, or due to rare variants or structural variants not included or difficult to capture with genotyping arrays [42]. Another possible reason for missing heritability is the lack of efficient fine-mapping the true disease-causing variant among the associated risk loci [142], that is that there might be rare variants with a greater effect size that are tagged by common SNP in the genotyping array showing a lower effect size when associated with disease, masking the true effect size of causal variants [142].

In summary, the unprecedented large and thorough phenotyped sample size enabled us to investigate with a genetic association study the common genetic variation underlying AAD, overcoming the main limitations met in previous genetic investigations. As such was the entire genome screened for association and both known and novel risk factors were identified. Future studies are warranted to link the associated genes to pathological mechanisms that can explain the origin of autoimmune Addison’s disease.

5.2 Paper II

By using the genotyping data from the GWAS in paper I, we developed a polygenic risk score (PRS14_AAD) as a tool to estimate the susceptibility to AAD at the individual level.

The ability of PRS14_AAD to discriminate between cases and controls was evaluated in an independent validation set. The area under the receiver operating characteristic curve (AUC) amounted to 0.88 (95% CI 0.87-0.90), which indicated a case-control discrimination in the range observed in other autoimmune conditions such as type I diabetes [143-146] and autoimmune thyroid disease [147, 148]. The average PRS14_AAD between cases and controls differed 1.5 standard deviations (p < 2e-16).
Given the low prevalence of AAD, PRS\textsubscript{14\textsubscript{AAD}} would not be suitable to predict disease development at individual level in the general population. Indeed, a random individual in the 90\textsuperscript{th} centile of PRS\textsubscript{14\textsubscript{AAD}} would have 0.4% risk of developing AAD.

Instead, we evaluated the negative predictive value of a low PRS\textsubscript{14\textsubscript{AAD}} in pediatric patients already diagnosed with AAD. As expected, we observed that most pediatric patients had a PRS\textsubscript{14\textsubscript{AAD}} above the average patient, suggesting an enrichment of risk loci in these cases that predisposed them to an earlier disease onset. In fact, the average PRS\textsubscript{14\textsubscript{AAD}} of cases that developed AAD at the age of 11 or earlier differed significantly from those that developed the disease at the age of 70 or later (p = 3e-4). Strikingly, two cases showed an exceptionally low PRS\textsubscript{14\textsubscript{AAD}} for pediatric AAD. We could confirm with WGS that both patients had non-autoimmune monogenic diseases, previously presumed to have AAD (Figure 3).

![Figure 3. PRS\textsubscript{14\textsubscript{AAD}} distribution in pediatric patients.](image)

**Figure 3. PRS\textsubscript{14\textsubscript{AAD}} distribution in pediatric patients.** Compared to controls and seropositive cases, the vast majority of pediatric patients had a PRS\textsubscript{14\textsubscript{AAD}} higher than the average case (red), suggesting an early onset due to an enrichment of risk alleles. In contrast, two patients presented a PRS\textsubscript{14\textsubscript{AAD}} lower than the average control (blue), indicating a possible monogenic rather than polygenic cause of disease. Indeed, whole-genome sequencing allowed us to confirm that patient 5 was homozygote for a disease-causing variant in the gene \textit{CYP11A1}, which leads to a rare form of congenital adrenal hyperplasia. Patient 10 was hemizygote for a novel insertion in the gene \textit{NR0B1}, which is known to cause adrenal hypoplasia congenita. Both patients were seronegative. Figure from Paper II [149].

Autoantibodies against 21-hydroxylase tend to slowly disappear over the years whereas PRS remains unchanged [72, 85-87, 150, 151]. Thus, if seronegative patients were truly autoimmune, they would display a similar average PRS\textsubscript{14\textsubscript{AAD}} as the seropositive subgroup. We found that patients lacking the hallmark autoantibodies against 21-hydroxylase had a significantly lower PRS\textsubscript{14\textsubscript{AAD}} compared to patients with AAD confirmed with autoantibodies (p = 6e-14). Seronegative patients recently diagnosed (less than 5 years), presented a significantly lower PRS\textsubscript{14\textsubscript{AAD}} compared to those with a
longer disease duration, here more than 20 years (p = 0.016). This result highlighted the possibility that mixed disease etiologies are present in the seronegative group.

Taken together, Paper II demonstrates a use case for the PRS14\textsubscript{AAD} harnessing the negative predictive value among patients diagnosed with primary adrenal insufficiency and indicates that some seronegative patients may have other disease etiologies than autoimmune.

### 5.3 Paper III

For Paper III, we decided to closely examine the seronegative subgroup. By selecting patients with low PRS among patients lacking 21-hydroxylase autoantibodies for WGS, we could diagnose 5 out of 35 (14%) with monogenic diseases, all of which were previously presumed to have AAD. We also identified variants of unknown significance that might be the true cause of primary adrenal insufficiency in three additional cases.

The majority of the re-diagnosed patients presented a later than expected age of onset for a monogenic form of primary adrenal insufficiency. Genetic defects that lead to monogenic Addison’s disease often present a wide phenotypic variability and age of onset. Hence, clinical manifestations of milder or non-classical forms can appear well in adulthood when detailed genetic inspection for primary adrenal insufficiency is not firstly considered during diagnostic workup [91, 106, 150].

In a heterogeneous condition as Addison’s disease, a personalized medicine approach is key to establish the correct etiology that enables adequate clinical management. The results of this study demonstrates the ability of clinical WGS in patients with idiopathic primary adrenal insufficiency and the potential role of PRS to select patients with most to benefit from diagnostic sequencing.
In this thesis, we overcame the sample size limitation of previous genetics studies in AAD by using the two largest DNA biobanks in the world with patients with this disease. This enabled us to perform a GWAS that could associate nine independent risk loci with the development of AAD. In particular, the strongest locus was the HLA region as previously reported, and the rest, were also implicated in different immune-related pathways. An uncommon coding variant in the AIRE gene involved in central tolerance mechanisms, was also strikingly associated. In future studies, functional assays may be useful to understand the biological role of associated genetic risk factors. The combination of a larger sample size and either genotyping or sequencing can contribute to uncover additional risk loci and explain additional missing heritability. A validation of these findings in an independent cohort may be the natural next step. Indeed, a candidate-gene study validated the associations of LPP and UBASH3A in a cohort from the UK [152].

We constructed a PRS14\textsubscript{AAD} that could estimate the disease susceptibility at individual level and helped to discriminate pediatric patients of different etiologies. Theoretically, if new risk loci were to be identified in a larger GWAS of the same ancestry, the accuracy of PRS14\textsubscript{AAD} would improve. Moreover, PRS14\textsubscript{AAD} holds promise as a tool to triage patients without 21-hydroxylase autoantibodies that may benefit from diagnostic sequencing, since 5 cases presumed to have autoimmune Addison’s disease could be diagnosed with WGS after stratifying seronegative patients by their genetic predisposition to AAD. More patients across the entire PRS14\textsubscript{AAD} range would be needed to sequence, in order to establish a PRS14\textsubscript{AAD} cut-off where the discrimination of different disease etiologies is most optimal. Thereby, a prompt molecular diagnose would be achievable.
This thesis was carried out at the Department of Medicine in Solna, in the Center for Molecular Medicine at Karolinska Institutet. Financial support was kindly provided by The Swedish Research Council, the Knut and Alice Wallenberg Foundation, Novo Nordisk Foundation, Stockholm County Council, KG Jebsen Foundation and the Research Council of Norway.

I am thankful to everyone who has contributed to this thesis, and specially:

I would like to express my most sincere gratitude to my main supervisor, Daniel Eriksson for being an outstanding supervisor. It has been an honor to be your first PhD student. Thank you for being reliable, generous and for unwavering support. For giving me the most realistic and honest opinions and your wise perspectives. Thanks for your patience and for always being available when needed, no matter how busy you were, for making the extremely complex simple and for sharing your knowledge and insights in research.

I would also like to sincerely thank my co-supervisor, Olle Kämpe, for welcoming me into your group and giving me the great privilege of doing my PhD studies in your laboratory, where many talented researchers have been trained. Thank you for always encouraging me to do my best, for believing in me from the very beginning and for providing me with countless smart advices and opportunities for learning, networking and developing me as a researcher. It has been a real pleasure to work in your group and enjoy of your endless source of knowledge, thanks for making me feel quickly integrated by teaching me what fika really means.

Thanks to my co-supervisor Sophie Bensing, for always giving me great input, smart advice and being so helpful and ahead of the curve with relevant documentation. Thank you for being the best bridge between the clinic and research, something I really have appreciated to better understand the translationality of our results.

Thanks to my co-supervisor Eystein Husebye for our research collaboration that made the projects presented in this thesis possible, for welcoming us to Bergen and for your great feedback. I extend this thanks to your group and Stefan Johansson, for the fruitful collaboration where I could learn a lot.

I would like to extend my gratitude to former and present members of Olle Kämpe’s and Sophie Bensing’s groups:

Nils Landegren, even though you were not officially one of my supervisors, I have learned a lot from you and am grateful for the opportunities you gave me to develop my skills as a researcher. I am particularly happy that we continue our collaboration now that you established your own research group, from which I had the pleasure of working with Axel Cederholm. It has been a pleasure to work with you Axel in the lab. Thank you for making the work in the lab so easy, for the good moments, the long experiments and for the creative and wild ideas. I am especially thankful that you accepted to be my toast master and, of course, for showing me the power of analog centrifuges. Åsa Hallgren,
thank you for being so welcoming when I started, for teaching me almost all I can do in the lab, for your good practices and your friendship. We miss you a lot and I am very happy you are back. Ileana Botusan, thanks for always having kind and encouraging words for me, I enjoy collaborating with you. Peri Noori, thanks for your time in the group, helping with our projects and for all your assistance when needed. Wafa Kamal, thanks for creating such a good environment in the lab when you are around, so sad that you are not a full-time PhD student. Fabian Sardh, I am happy to work with you again, thank you for your support during the writing of this thesis. Jakob Skov, Sara Sandersjöö-Fletcher, Sara Öster, I am happy to have met and worked with you, thanks for all the nice moments.

Thanks to all the collaborators at the Department of Endocrinology and the Swedish Addison Registry, Per, Anna-Karin, Olov, Ragnhildur, Magnus, Jeanette. Thank you to Cindy Wong, for excellent review of our manuscripts.

I further would like to express my gratitude to the great researchers and new friends I have met along the way of this PhD:

Ola Nilsson, thank you for your smart input and tricky questions. Lars Ottoson, thanks for always being so kind and helpful and for giving me quick master classes in lab management best practices when most needed. Vijole Ottoson, Marta Baroncelli, Ameya Bendre and Zelong Dou, thanks for enjoyable conversations and good times in and out of the lab. Agneta Laurent, thanks for being so caring and the best late-evening corridor mate I could have. Sho, thanks for kindly checking up on me during my writing. Lina Diaz, thanks for your realistic points of view, your generosity, your endless source of energy and motivation and for welcoming me as a monitor in your doctoral course. Ana, we met recently but quickly felt like we had known each other for a long time; Mike, you always find the right questions to ask; thanks to both of you for your generosity, the fun times, the good food and the right amount of projectiles. Jan, we (re)met pretty late in the studies, but better late than never. Thanks for your kind words and reminding me about the life behind lab’s door. All in all, Ana, Jan and Anita, thanks for being great ‘coworkers’. Anthonie, thanks for being so open to talk about any kind of topic, from T-cells and clustering to Swedish grammar and, the best of the topics, minoritetsspråk. Thank you for being so supportive and always sharing your opinions honestly.

I could not forget an important part of this PhD, the National Clinical Research School in Chronic Inflammatory Diseases. Many thanks to Helena Erlandsson Harris and Cecilia Aulin for welcoming me to the school and for your efforts in organizing excellent workshops. Special thanks to Cecilia for kindly accepting to be the chairperson at my dissertation. Thanks to all the fantastic colleagues I met at the research school: Anna, Charlotte, Eveline, Fabiola, Heshuang, Rebecka, Roberta, Maike, Qiang, Remigius, Samuel, Sebastian, Soran, Tilen, Ian, Xinxin, for creating a nice and knowledge-rich environment. Special thanks to Xinran and Yunbing for your friendship.

Special thanks to my closest friends: Irina, thank you for your friendship, your generosity, for being reliable and, of course, for convincing Sophie of our productive teamwork. Yitian and Liuting, thanks for your support and nice times together. Dave, thanks for your encouraging words, for your special smoothies and for keeping in touch despite the distance. Raquel, Laia and Benjamin, thanks for all the fun at Swedish courses and afterwards. Somehow we managed to learn the language even
though I do not remember ever speaking Swedish between us in class. *Marina, Melina, Maria, Joan* and *Jorge*, thanks for making possible to keep our friendship despite the time and distance.

Lastly, I would like to express my deepest gratitude to my dearest. *Ana*, thanks for unmatched and unconditional support, for your encouraging words, your wise advices and for showing us that we could do anything we wanted. *Adrián* and *Carlos*, the best brothers I could have ever asked for. Thanks for always being there at the right time, with the right comment, criticism and support, and for being the only ones able to make me laugh until cry. *My parents*, thanks for your unfailing support, for respecting and understanding my decisions, for your countless sacrifices to always give us the best, for your wise advice and for providing us the best of the educations.


