Genetic Variation and Expression of the IRF5 Gene in Autoimmune Diseases

GUDLAUG THORA KRISTJANSDOTTIR

Dissertation presented at Uppsala University to be publicly examined in Enghoffsalen, Akademiska sjukhuset, ingång 50, bv, Thursday, April 16, 2009 at 09:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

Abstract
Kristjansdottir, G T. 2009. Genetic Variation and Expression of the IRF5 Gene in Autoimmune Diseases. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 434. 51 pp. Uppsala. ISBN 978-91-554-7450-8.

The interferon regulatory factor 5 (IRF5) gene encodes a transcription factor that plays an important role in the innate as well as in the cell-mediated immune response. The IRF5 gene has received considerable attention since it was shown to be associated with two autoimmune diseases; systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA). The aim of this thesis was to examine if *IRF5* is associated with other autoimmune diseases and to investigate the role of the genetic variation of *IRF5*. In the first study a set of common polymorphisms in *IRF5* were analyzed for their association with two subgroup of inflammatory bowel disease (IBD); Crohn´s diseases (CD) and ulcerative colitis (UC). A strong signal of association of *IRF5* with IBD was found. The most strongly associated polymorphism is a 5 base pair (bp) insertion-deletion (indel) in the promoter region of the IRF5 gene. The association was detected within both UC and CD, and appeared to be stronger in UC. In the second study we investigated the association of *IRF5* with multiple sclerosis (MS). A similar set of polymorphisms as in the IBD study were genotyped in a cohort of MS patients and controls. The same polymorphisms that were associated with IBD were also found to be associated with MS. In the third study, we performed a comprehensive investigation of the IRF5 gene to detect most of the polymorphisms in the gene, and to determine to what extent they account for the association signals obtained from the gene. *IRF5* was sequenced and 34 new polymorphisms were identified. Twenty seven of these, and 20 previously known SNPs in *IRF5* were genotyped in an SLE case-control cohort. We found that only two polymorphisms, the 5bp indel and a SNP downstream of *IRF5*, account for the association signal from all the remaining markers in the IRF5 gene, and that these two polymorphisms are independently associated with SLE. Interestingly, in our studies on IBD and MS, we only observed the signal from the 5bp indel polymorphism as a risk factor for IBDs. In the fourth study the two independent risk alleles in *IRF5*, were tested for their association with primary Sjögren´s syndrome (pSS). In this study we also included one SNP in the STAT4 gene, since *STAT4* had recently been shown to be associated with SLE. Both risk factors in *IRF5* and *STAT4* were found to be associated with pSS.

The regulation of expression of *IRF5* was also investigated in the first three studies. We observed allele-specific differences in protein binding as well as increased binding of the transcription factor SP1 to the 5bp risk allele. We also detected increased expression of the *IRF5* mRNA from a promoter containing the risk allele.

Taken together, the results of our studies suggest a general function for *IRF5* as a regulator of the autoimmune response, where the 5bp indel is associated with IBD, MS, SLE and pSS. The additionally polymorphisms, which account for the remaining association signal obtained with SLE and pSS, may contribute to the disease manifestations that are specific for rheumatic diseases. Our studies add to the evidence that there are genes or pathways that are common in multiple autoimmune diseases, and that the type I interferon system is likely to be involved in the development of these diseases.

Gudlaug Thora Kristjansdottir, Department of Medical Sciences, Molecular Medicine, Uppsala University, SE-75185 Uppsala, Sweden

To my children,
Guðrún, Þorvarður and Kristján

List of papers

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

I   Dideberg, V, **Kristjansdottir, G**, Milani, L, Libioulle, C, Sigurdsson, S, Louis, E, Wiman, A C, Vermeire, S, Rutgeerts, P, Belaiche, J, Franchimont, D, Van Gossum, A, Bours, V, Syvanen, A C. <u>An insertion deletion polymorphism in the Interferon Regulatory Factor 5 (IRF5) gene confers risk of inflammatory bowel diseases</u> Hum Mol Genet. (2007) 16(24) 3008-16. Epub 2007 Sep 19.

II **Gudlaug Kristjansdottir***, Johanna K. Sandling*, Alessandro Bonetti, Izaura M. Roos, Lili Milani, Chuan Wang , Sigrun Gustafsdottir, Snaevar Sigurdsson, Anders Lundmark , Pentti J. Tienari, Keijo Koivisto, Irina Elovaara, Tuula Pirttilä, Mauri Reunanen, Leena Peltonen, Janna Saarela, Jan Hillert, Tomas Olsson, Ulf Landegren, Antonio Alcina, Oscar Fernández, Laura Leyva, Miguel Guerrero, Miguel Lucas, Guillermo Izquierdo, Fuencisla Matesanz, and Ann-Christine Syvänen. <u>Interferon Regulatory Factor 5 (IRF5) Gene Variants are Associated with Multiple Sclerosis in Three Distinct Populations</u> J Med Genet. (2008) 45(6) 362-9. Epub 2008 Feb 19.

III Sigurdsson, S, Goring, H H, **Kristjansdottir, G,** Milani, L, Nordmark, G, Sandling, J, Eloranta, M L, Feng, D, Sangster-Guity, N, Gunnarsson, I, Svenungsson, E, Sturfelt, G, Jonsen, A, Truedsson, L, Barnes, B J, Alm, G, Ronnblom, L, Syvanen, A C. <u>Comprehensive Evaluation of the Genetic Variants of Interferon Regulatory Factor 5 Reveals a Novel 5bp Length Polymorphism as Strong Risk Factor for Systemic Lupus Erythematosus</u> Hum Mol Genet. (2008) 17(6) 872-81. Epub 2007 Dec 6.

IV Gunnel Nordmark[*], **Gudlaug Kristjansdottir**[*], Elke Theander, Per Eriksson, Johan G Brun, Chuan Wang, Leonid Padyukov, Lennart Truedsson, Gunnar Alm, Maija-Leena Eloranta, Roland Jonsson, Lars Rönnblom and Ann-Christine Syvänen. <u>Additive effects of the major risk alleles of IRF5 and STAT4 in primary Sjögren's Syndrome</u> Genes Immun. (2009) 10(1) 68-76. Epub 2008 Dec 18.

These authors contributed equally to the work.

Supervisor
Professor Ann-Christine Syvänen
Department of Medical Sciences, Molecular Medicine
Uppsala University, Sweden

Co-supervisors
Professor Lars Rönnblom
Department of Medical Sciences, Rheumatology
Uppsala University, Sweden

Dr Gunnel Nordmark
Department of Medical Sciences, Rheumatology
Uppsala University, Sweden

Faculty opponent
Dr Unnur Þorsteinsdóttir
deCODE Genetics
Reykjavík, Iceland

Review board
Professor Lars Hellman
Department of Cell and Molecular Biology
Biomedical Center, Uppsala, Sweden

Docent Marie Carlson
Department of Medical Sciences, Gastroenterology and hepatology
Uppsala University, Sweden

Dr Kristina Duvefelt
Clinical Research Centre, Mutation Analysis Facility
Karolinska University Hospital, Huddinge, Sweden

Contents

Abbreviations

ANA	Antinuclear antibody
ddNTP	Dideoxynucleotide triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
CD	Crohn´s disease
FP	Fluorescence polarization
GWA	Genome wide association
HapMap	The haplotype mapping project
HLA	Human leukocyte antigen
IBD	Inflammatory bowel disease
IFN	Interferon
Indel	Insertion/deletion
IRF	Interferon regulatory factor
ISRE	Interferon stimulated response element
LD	Linkage disequilibrium
MHC	Major histocompatibility complex
MS	Multiple sclerosis
NIPC	Natural interferon producing cells
OR	Odds ratio
PCR	Polymerase chain reaction
PDC	Plasmacytoid dentritic cells
pSS	Primary Sjögren´s syndrome
RNA	Ribonucleic acid
SSA/Ro	Sjögren´s syndrome antigen A
SSB/La	Sjögren´s syndrome antigen B
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
sSS	Secondary Sjögren´s syndrome
TLR	Toll like receptor
UC	Ulcerative colitis
UTR	Untranslated region

Introduction

Modern genetics began with the work of an Austrian monk, Gregor Mendel during the mid 19th century. Mendel observed that organisms inherit traits in a discrete manner following distinct laws. These laws were later named after him, the Mendelian laws of inheritance. The biological basis for inheritance remained more or less unknown until 1944, when DNA was found to be the carrier of genetic information[1]. The characterization of DNAs structure as a double helix, and the explanation how this double helix both contains and transmit genetic information was achieved nearly ten years later, by James Watson and Francis Crick[2; 3]. Since then, there has been a remarkable achievement in the understanding of genetic variation. The reference sequence of the human genome was released in April 2003, when the International Human Genome Sequencing Consortium announced the successful completion of the Human Genome Project[4]. In September 2007, about four years after the human genome sequence was published, the complete six-billion-bases genome of an individual human was unveiled for the first time. This was the genome of J. Craig Venter[5], former president and founder of Celera Genomics. The field of individual genotyping continues to expand with an enormous speed. In January 2008, a project called "1000 Genomes Project" (www.1000genomes.org) was established. The scientific goal of this project is to sequence the human genomes of over 1000 individuals and by that map all the variants found in the human genome. The overall aim of the project is to provide a detailed and medically useful map of human genetic variation.

Genetic variation

It has been confirmed by the whole genome sequencing projects, that any two unrelated people are more than 99 percent the same at the nucleotide level.

Most of the genetic differences that vary among people do not have any effect on the individual, although some may explain individual differences in disease susceptibility and response to environmental factors. The effort to find and understand these genetic differences is therefore both an exciting and an important task.

Genetic polymorphisms include single nucleotide polymorphisms (SNPs), insertion-deletion polymorphisms (indels), repetitive sequences, copy number polymorphisms and other complex forms of DNA variation.

Single nucleotide polymorphisms

The most common form of human genetic variation are the SNPs, that occur in about 1:1000 bases in the human DNA. This type of variation is often defined as a position in the genome where a single nucleotide differs between individuals and that this change is seen in at least 1% of the population. The SNP itself can either directly affect the function of a gene or be useful as a marker in search for such functional variation. The ratio of non-synonymous SNPs to synonymous SNPs in the coding region of an individual genome has been estimated to be nearly 1:1. The majority of the non-synonymous SNPs are believed to be functionally neutral[6].

Currently more than 14.7 million SNPs are registered in NCBI's SNP database and about half of them are validated (www.ncbi.nlm.nih.gov/SNP; build 129). The number of verified SNPs available in freely accessible public databases is increasing, mainly because of the efforts of the international HapMap project (www.hapmap.org), which was established in 2002.

Insertions/Deletions

The second most common form of human genetic variation are the indels[7]. It has been predicted that indels represent between 16% and 25% of all sequence polymorphisms in humans[8]. Most indels are short, only 1-6 nucleotides in length. Enrichment of indels with a size that is a multiple of three, have been observed in exons. Indels have been found to be underrepresented in coding regions compared to the entire genome, and in protein-coding regions they are typically located in regions that encode the N- or C-terminus[6].

Indels have been shown to be in linkage disequilibrium (LD) with common SNPs, and to have a similar population distribution[7]. They can, in the same way as SNPs, be used as markers in genetic studies or even be a functional variant of a gene.

Repetitive sequences

Repetitive sequences are grouped into two main types, depending of their occurrence in the DNA; tandem repeats which are repeated directly following each other, and interspersed repeats which are found across the genome. The interspersed repeats are further divided into two classes depending on their size, SINES (Short Interspersed Elements) less than 500bp long and LINES (Long Interspersed Elements) more than 500bp in length.

The tandem repeats are divided into three classes depending on the length of the repeated unit; satellites with repeats of several kb in length, minisatellites with repeats between 10-100bp and microsatellites with repeats of only 1-6bp in length. The microsatellites are commonly used as genetic markers especially in linkage studies and in forensic genetics.

Copy number variation

It has only recently been discovered that copy number variations (CNVs) are widespread and a common form of variation in the human genome. CNVs are DNA segments that are more than 1kb in length and for differences in copy number have been observed by comparison of two or more genomes[9]. CNVs segments can be up to hundreds of kb in size and they are often found to be in tight LD with nearby SNPs[10]. They can be caused by genomic rearrangements such as deletions, duplications, inversions, and translocations.

Genetic diseases

Human inherited disorders can be classified into four categories base on their genetic complexity: monogenic, mitochondrial, X-linked and complex diseases. To date, over 2200 genetic disorders are known[11].

Monogenic

Monogenic diseases are a group of genetic disorder, caused by a change affecting one single gene. These diseases follow the pattern of Mendelian inheritance, either autosomal or X-linked. Most known single gene diseases follow the autosomal pattern, meaning that the disease causing gene is located on a chromosome that is not a sex chromosome. Both autosomal and X-linked disorders are inherited either in a recessive or dominant manner.

A dominant disorder is phenotypically expressed in a person who has only one copy of the disease-causing allele (heterozygote). Neurofibromatosis is an example of an autosomal dominant disease[12].

An individual suffering from a recessive disorder has two copies of the disease allele (homozygote), to express the disease phenotype. Cystic fibrosis is an example of autosomal recessive disorder[13]. An example of X-linked recessive disorder is red-green color blindness[14]. This disease is much more common in males than females, since two copies of the disease allele are required for the disease to occur in females, while only one copy is required in males. Fragile X is an example of an X-linked dominant disease[15].

Mitochondrial diseases

In mitochondrial disorders, the disease causing gene is located in the mitochondrial genome. The mitochondrial DNA is inherited solely from the mother, and this type of disease is therefore maternal inherited. An example of mitochondrial diseases is Leber's hereditary optic neuropathy[16].

Complex diseases

Complex diseases are called complex, since they do not display the simple Mendelian pattern of inheritance. These diseases are caused by multiple

genetic, environmental and lifestyle risk factors. The proportion of genetic and environmental risk factors varies between diseases. Disease concordance and allele sharing among relatives can be used to separate genetic and environmental factors. Twin pair studies are often used for this purpose. Twins are assumed to be exposed to environmental risk factors to the same extent whether they are monozygotic (MZ) or dizygotic (DZ). If there is greater concordance in MZ than in DZ twins, it is seen as an evidence of a genetic component of the disease.

The vast majority of known diseases are complex diseases. Examples of complex diseases are all the autoimmune diseases studied in this thesis.

Identifying human disease causing genes

So far, the major impact of genetic research in medicine has been by defining genes causing monogenic diseases. However during the last few years, numerous variants contributing to the formation of complex diseases have been identified. This is in many ways the consequence of the completion of the human genome sequencing project, and the development of rapid and relatively cheap genotyping techniques.

Two major approaches are used for identifying genes underlying genetic diseases, linkage studies and association studies.

Linkage studies

Linkage analysis usually involves genotyping genetic markers (often microsatellites) across the genome, in large families, where the disease affects individuals in several generations. The goal is to find markers that co-segregate with the disease. The basis for the cosegregation is the tendency of alleles at adjacent loci on the same chromosome to be inherited together without recombination, as a consequence of their physical proximity. Linkage is calculated for each marker and expressed as the logarithm of the odds (LOD), of two markers being linked compared with independent of each other. The significance threshold for linkage is traditionally considered to be a LOD score greater than 3.0. Linkage analysis can be used both in mapping Mendelian diseases (parametric linkage) and complex diseases (model-free or non-parametric linkage).

The main limitation of linkage studies is their low resolution, since the analysis is limited to the number of meioses in the family material used. Therefore a disease locus found by linkage analysis is usually large, up to several Mb in size and can contain many candidate genes.

A large number of genes causing monogenic diseases has been found by linkage analysis. An example is the transcriptional enhancer TEAD1 protein,

causing Helicoid peripapillary chorioretinal degeneration (HPCD) [17], mono-genic eye disease with high penetrance.

Association studies

In association studies, the co-occurrence of certain alleles of a polymorphism (often defined by a SNP) and a phenotypic trait is calculated.

Genetic association studies can either have family-based design or be case/control studies. A commonly used family based association test is the transmission disequilibrium test (TDT), which we used in Study II in this thesis. This test measures the transmission of alleles from heterozygote parents to an affected offspring. If the marker is associated with the disease, the allele conferring increased risk will be transmitted to the affected offspring more often than expected by chance.

In case/control studies, the frequencies of alleles are compared between two groups, patients and controls. A significant difference in the allele frequency between the two groups indicates that the genetic marker is associated with the disease, or is in linkage disequilibrium (LD) with a polymorphism that does. When markers are said to be in LD, it means that they occur together more frequently in the population than they would do, if they were combined at random. LD between markers is often described using the terms D´ or r^2, ranging from 0 to 1.

D´=1, describes complete linkage between loci, no recombination has taken place between the markers and they are therefore inherited together. This however, does not provide any information about the frequency of the alleles in the population analyzed, which can be the same for the loci or different. r^2 on the other hand, describes the correlation of allele frequencies between the loci. When $r^2=1$, the alleles are in complete linkage and have the same frequency in the sample analyzed. This means, that by knowing the genotype of one marker, the genotype of the other is also known. This correlation is used to select informative SNPs, also called Tag-SNPs, for genotyping in large association studies.

There are two approaches when performing association studies, either to search for association with specific candidate genes or to perform a genome wide association studies with no prior hypothesis of the localization of the susceptibility variants.

Genome-wide association studies

Genome wide association studies (GWAS) are large scale association studies where several hundred thousand, up to a million SNPs, located across the genome are genotyped in large sample cohorts. It has recently become possible to perform association studies on this scale, due to modern genotyping techniques, together with freely available SNP databases containing information about the LD pattern in the human genome [18].

The huge number of genotyped markers increases the risk of obtaining false association signals by chance. Therefore a correction for multiple testing is needed in GWAS. A threshold of 1×10^{-8} have been suggested for declaring a significant association in GWAS[19]. This is equivalent to a p-value of 0.05 after Bonferroni correction for 1 million independent tests. The Bonferroni correction is widely used to correct for in multiple testing. In this correction it is assumed that the tests are independent of each other, which is often not true in genetic studies, since many of the markers are in LD and therefore not independent. In Study II, we used the Bonferroni method to correct for multiple testing of haplotype associations with MS. The Bonferroni test can be considered as too conservative and as such increases the risk of rejecting true findings. Permutation testing is less stringent than Bonferroni as it takes into account the LD between markers.

Since the first publication of a GWAS in March 2005[20], convincing positive results are regularly being published. More than 300 replicated associations have been reported for more than 70 complex diseases[21]. GWAS have been performed for three of the diseases studied in this thesis. Until now, eleven GWAS have been published for IBD[22-32], five for MS[33-37] and five for SLE[38-42]. To my knowledge, no genome wide association study has been performed in Sjögren´s syndrome.

Autoimmune diseases

The autoimmune diseases are a large group of diseases, with a great variation in their clinical manifestations. They are complex diseases, where both genes and environment contribute to disease development. In all of these diseases the underlying problem is similar, inappropriate immune reactivity against self antigens. The breaking of tolerance may result in both self-reactive B- and T-lymphocytes, and the effect can be either organ specific or systemic. Autoimmune reaction to antigens found only in certain tissues or organs will result in a localized disease process, while autoreactivity against ubiquitously expressed antigens may results in systemic autoimmune disease. Inflammatory bowel diseases (IBD) and multiple sclerosis (MS) are examples of organ-specific autoimmune diseases, where the immune-mediated injuries are localized to the gastrointestinal tract and the central nervous system, respectively. In contrast, systemic lupus erythematosus (SLE) and Sjögren´s syndrome (SS) are examples of non-organ-specific diseases, since the immune reaction is directed against many different organs and tissues, resulting in widespread inflammation. Autoimmune diseases are often thought to be the third largest category of illnesses in the industrial world, after cancer and cardiovascular diseases. Approximately three to five in 100 individuals suffer from a recognized autoimmune disorder[43; 44], although this number varies depending on what diseases are taken into ac-

14

count. The common autoimmune diseases, such as RA and thyroiditis are estimated to affect about 0.5% of the human population[45]. For reasons that are not clear, the prevalence of autoimmune diseases appears to be rising. They are a particular threat to women, especially during the fertile years. The gender difference in incidence decreases among older patients. Some reports suggest differences in the rates of autoimmune disease among various ethnic groups. An example of this is in the United States, were African Americans are at higher risk than Caucasians for having SLE [46], but at lower risk for having IBD[47] and MS[48].

Shared genes in autoimmune diseases

It is relatively common that patients affected by an autoimmune disease suffer from more than one autoimmune disease, and that members of the same family suffer from different autoimmune diseases. Such observations suggest shared genes or involvement of common cellular pathways in these diseases. This hypothesis has been widely studied, and it has been predicted that most of the genetic predisposition to autoimmune diseases can be explained by genes residing in the major histocompatibility complex (MHC)-locus[49]. Although genes outside the MHC region, harboring susceptibility to multiple autoimmune diseases, have also been reported, such as *PTPN22* in RA, SLE, celiac disease, Graves disease and psoriasis[50-53], *CD24* in MS and SLE[54], *TNFAIP3* in RA, SLE and type I diabetes[41; 55; 56] and *IRF5* in SLE, RA and SS[57-59].

By investigating gene expression profiles in the peripheral blood mononuclear cells (PBMCs) of patients with autoimmune diseases, a common signature of gene expression in patients with RA, SLE and MS has been described[60]. On the other hand, some genes have been found to be highly associated with a particular autoimmune disease, but not with others. This suggests that there are at least two types of autoimmune susceptibility loci; those that are shared among autoimmune diseases and those that are specific for an individual disease.

Inflammatory bowel diseases

Inflammatory bowel diseases (IBDs) are common causes of gastrointestinal morbidity in Western countries with a prevalence of 40 per 100,000 persons. The highest incidence rates are reported in Northern and Western Europe and in North America, while the lowest incidences are found in Africa, Asia and America[61]. The IBDs are divided into two diseases, Crohn´s disease (CD) and ulcerative colitis (UC), which are distinguished by the localization of the intestinal inflammation and the pathology of the inflamed tissue. In CD the inflammation can involve any portion of the gastrointestinal tract, while in UC it is localized to the colon, and the patients usually present with bloody diarrhea[62]. Although the pathogenesis of IBD is uncertain, the role of genetic factors in

IBD is supported by a strong familial clustering of the disease and by higher disease concordance between monozygotic (MZ) than between dizygotic (DZ) twins [63]. Of the two phenotype groups of IBD, CD and UC, the underlying genetics of CD has been studied in much more detail.

An effect on the predisposition to CD has been demonstrated and validated for a few genes, e.g. *NOD2*[64; 65], *IBD5*[66; 67] and *ATG16L*[68]. Six genome wide association studies have been published, confirming the previously known associated genes, and bringing the number of known independent CD associated loci to more than 30[22-27]. The variations so far identified by the GWAS in CD have been estimated to account for 10% of the inherited risk of the disease[11].

Two genome wide association scans have been performed in UC. Beside the MHC region and the IL23R gene, which has previously been identified both in CD and UC, three new loci where found on 1p36, 12p15 and 1q32 where, the IL10 gene is located[28; 29]. Additionally two new loci on 20q13, 21q22, respectively and NELL1 has been identified as a novel IBD disease gene[30; 32].

Multiple sclerosis

Multiple Sclerosis (MS) is estimated to affect over 2 million individuals worldwide. The frequency of the disease varies greatly depending on population and geographic location. It is most common in Caucasians of northern European descent, with a frequency of about 1:1000 inhabitants[69]. The highest prevalence of MS is reported from the Orkney Islands of Scotland, with >250 cases per 100,000 inhabitants. This high prevalence may be partly due to intensive prolonged search for MS in this small population [70]. MS is not well recognized as an autoimmune disease, but without doubt MS exhibits features of autoimmunity. The disease shows familial clustering, and twin studies have revealed that a large portion of this clustering can be attributed to shared genes[71; 72]. Despite on-going efforts, a remarkably low number of susceptibility genes for MS have been identified. Genome-wide linkage studies have indicated several possible susceptibility loci, but the only locus to be identified across most of these studies is the MHC-locus on chromosome 6p21, were the HLA-DRB1*1501 allele is a well established genetic risk factor for MS[73; 74]. Numerous candidate gene studies have also been performed in MS, although findings from one population have been difficult to replicate in other populations. It is only recently that genes residing outside the MHC have been found to be associated with MS, in more than one population. Protein kinase C alpha (PRKCA), is one of these few genes[75; 76]. Interleukin 7 receptor alpha chain gene (IL7RA) and interleukin 2 receptor alpha chain gene (IL2RA) were found to be associated with MS in independent candidate gene studies and in a genome-wide association study[33; 77-79]. These finding have already been replicated in other populations[80].

Systemic lupus erythematosus

The prevalence of systemic lupus erythematosus (SLE) has been reported to be 20-68 cases per 100,000[81], with a very high gender difference, where nine out of ten patients are females. SLE is a chronic disease, characterized by multiple organ involvement and production of autoantibodies, which together with their corresponding antigens form immune complexes (IC). These IC are thought to be the major cause of inflammation in SLE.

SLE is classified according to the ACR criteria[82] but a clinical diagnose of SLE is typically made when two organ manifestations together with a characteristic SLE serology are identified.

There is a well documented familial segregation of SLE, and twin pair studies reveal about a 10-fold higher rate of concordance for SLE for MZ twins compared with DZ pairs[83]. The genetics of SLE has been intensively studied since SLE is often thought of as a prototype for autoimmune diseases. Numerous genes have shown to have consistent association with SLE even between ethnic groups e.g. Fcγ receptors[84],*PDCD1*[85],*PTPN22*[52],*IRF5*[57] and genes in the highly polymorphic MHC-locus. Recently four genome wide scans in SLE have been published[39-42]. In these studies several chromosomal regions that potentially harbor SLE-susceptibility genes and many candidate genes reached genome significance. Strong association was found with the HLA locus and the previously known SLE genes IRF5 and STAT4. Additionally three new risk variants were found, *C8orf13-BLK*, *ITGAM*-ITGAX and *BANK1*.

Sjögren´s syndrome

Sjögren´s syndrome (SS) is one of the most common autoimmune disease with an estimated prevalence of around 0.5% in Caucasian women[86]. The gender difference is similar to that of SLE, with a female: male ratio of 9 to 1[87]. It is a chronic disease characterized by lymphocytic infiltration of the salivary and lacrimal glands, causing significantly decrease in the quantity of saliva and tears. The hallmark of the disease is dryness in both mouth and eyes, although the disease can involve other organs, such as kidneys, gastrointestinal tract or lungs. Sjögren´s syndrome is classified according to the American-European consensus criteria (AECC)[88] and can be divided into primary and secondary. Primary Sjögren´s syndrome (pSS) patients do not have any other recognized autoimmune rheumatic disease, while secondary Sjögren´s syndrome (sSS) patients suffer from another rheumatic disease, often RA, SLE or scleroderma. The etiology of Sjögren´s syndrome is largely unknown. No population based studies of familial aggregation of SS or comparison of concordance between MZ and DZ twins have been performed[89]. Despite the lack of publication about the inheritance of SS, there is no doubt that a genetic component contributing to the disease development. Relatively few genetic studies have been

published and few genes have been shown to be associated with the disease; *IL-6* and *IL-10*[90; 91], the FAS-receptor (*CD95*) and FAS ligand (*FasL*)[89], *IRF5*[58] and most recently *STAT4*[92].

The interferon system

The term interferon (IFN) was first used in 1957, by Isaacs and Lindenmann, when they described a protein that induced antiviral resistance in mammalian cells [93]. Today we know that there exist several different IFNs and they have been classified into type I, II and III, based on amino acid sequence homology. The type I IFN gene family is located on chromosome 9, the genes are intron-free and show strong homology in structure [94]. Type I interferon consist of eight classes (α, β, ε, δ, κ, τ, ω, and limitin). Two of these classes, IFN-α and IFN-β, are the most studied. The human genome contains at least 20 different IFN-α genes, of which 13 are functional, and only a single known INF-β gene. The type II IFN, also known as IFN-γ, is encoded by a gene located on chromosome 12. The most recently described group of IFNs are the type III IFNs with three members termed IFN-λ1, IFN-λ2 and IFN-λ3, located on chromosome 19[95].

The type I IFNs, have received considerable attention since it was shown that treatment with IFN-α could induce autoimmune diseases[96]. Increased expression of IFN-induced genes has been detected in several autoimmune diseases. This so called IFN-signature has been identified in several of the diseases studied in this thesis, SLE[97], RA[98], Sjögren's syndrome[99] and in a subgroup of MS patients[100]. Type I IFNs have a large number of different effects on the immune system. In addition to the classical antiviral activity, type I IFNs also contribute to the antiviral immune response by stimulating the cytotoxic activity of natural killer cells, maturation of dendritic cells (DCs), and promotion of various T-cell functions[101]. Recombinant or purified type I IFNs from leukocytes are used as drugs for a variety of conditions. In chronic virus infections IFN-α is a common treatment, and IFN-β is used in multiple sclerosis at the time of diagnosis[102]. Although IFN treatment can be helpful, virtually all patients receiving IFN therapy experience some kind of side effect. Common adverse effects are a flu-like syndrome with headache and fever. There are reports of patients developing severe autoimmune manifestations, including SLE after receiving IFN-α treatment[103; 104].

Type I IFN production

Most nucleated cells can produce type I IFNs under certain conditions, though the major type I IFN-producing cells are the plasmacytoid dendritic cells (PDCs) also called the natural interferon producing cells (NIPC)[105]. This cell type is rare, but produces large amount of IFN-α per cell (1-2U/cell). The

PDC are found in the circulation of healthy individuals, but are mainly re-
cruited to lymphoid and other organs, e.g. skin in SLE patients [106]. The pro-
duction of type I IFNs are mainly triggered through the Toll-like receptors
(TLRs), which are highly conserved receptors, even between species[107]. Ten
members of the human TLRs have been identified [108]. TLR7 and 9 are ex-
pressed in PDCs, and are found in the endoplasmic compartments. TLRs
recognize so called PAMPs (pathogen associated molecular patterns), which
are small molecular motifs consistently found on pathogens. Binding of
PAMPS to TLRs, which in the case of TLR7 or 9 are single-stranded RNA or
CpG rich DNA motifs[109], trigger a signaling pathway, leading to activation of
the interferon regulatory factors (IRFs) 3, 5 and 7. The IRFs are latent cyto-
solic proteins that are activated upon phosphorylation[101]. The phosphorylated
IRFs dimerize and translocate to the nucleus, where they activate transcrip-
tion of type I IFN genes. In addition to this exogenous triggering, type I IFN
production can also be induced by immune complexes (IC), consisting of
nucleic acids and autoantibodies. This mechanism has been shown both in
SLE and SS patients. IC binds to FcγRIIa on the cell surface of PDCs, leading
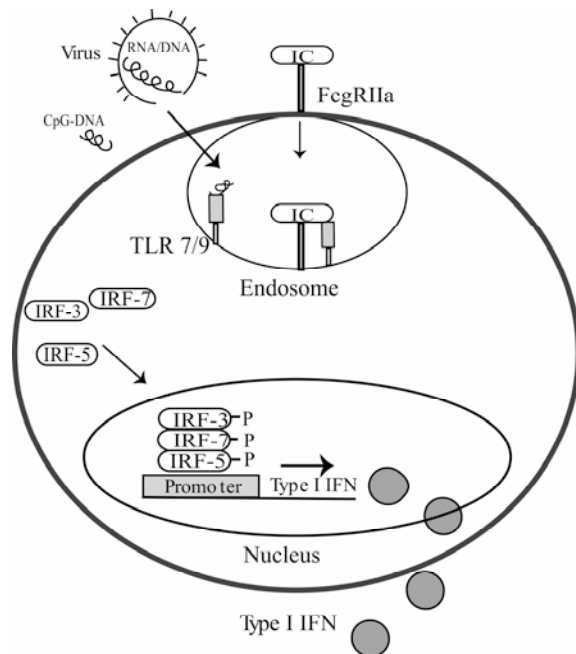to endosomal uptake of the complex[110; 111]. In the endosome the IC- FcγRIIa

Figure 1. Type I IFN production in plasmacytoid dendritic cells. Activation is in-
itiated, either by single stranded RNA or CpG-rich DNA binding to TLR7 and 9 in
endosomes, or by immune complexes binding to FcγRIIa on the cell membrane. The
activation leads to phosphorylation of IRF3, 5 and 7, turning on transcription from
type I IFN genes.

19

complex cross binds to the TLRs (Figure 1). This mechanism for type I IFN production has been shown in vitro for autoantibodies, from both SLE and SS patients[112; 113].

The type I IFNs produced are secreted from the cell and can bind to the type I IFN receptors (IFNAR), which are expressed on the cell membrane of most cell types, including PDCs. The binding initiates a signaling cascade in the cytosol, leading to activation of transcription of the IFN inducible genes.

Interferon regulatory factors (IRFs)

The role of the interferon regulatory factors (IRFs) in the regulation of the immune system and oncogenesis is well established. The family of IRF proteins consists of nine members, IRF1-9[114]. All known human IRFs are similar in structure. They possess a conserved helix-turn-helix DNA binding domain (DBD), of approximately 120 amino acids. This DBD recognizes the interferon stimulated response element (ISRE), located at the promoter region of IFN inducible genes. Notably, the promoter regions of all known IRFs contain this same conserved ISRE motif, indicating that IRFs themselves are regulated by type I IFN.

For the scope of this thesis, IRF5 is the most interesting one. The human IRF5 was originally characterized for its role in virus-induced type I IFN signaling[115]. It has since then been shown to be involved in the innate immune response to pathogens, by activating transcription of not only type I IFNs, but also proinflammatory cytokines, e.g. interleukin(IL)-6, IL-12 and TNF-α[116; 117]. IRF5 has also been shown to have tumor suppressor activity and to stimulate apoptosis in virally affected, as well as DNA damaged cells[117; 118]. The IRF5 gene displays a complex transcription pattern and at least nine alternatively spliced mRNAs have been described[119].

Present studies

The Interferon regulatory factor 5 (IRF5) gene encodes a transcription factor that plays an important role in both the innate and the cell-mediated immune response. When I started this project, the IRF5 gene had recently been shown to be associated with SLE and RA by our group, in collaboration with the group for systemic autoimmunity here in Uppsala.

Since then, these findings have been convincingly replicated in numerous populations. Despite the convincing replication of the association between *IRF5* and SLE in several populations, it was not apparent which the associated functional polymorphisms in the IRF5 gene were.

Taking into consideration that different autoimmune diseases might share common pathological pathways, we studied whether the IRF5 gene would be involved in susceptibility to other autoimmune diseases than SLE, and what the possible role of the genetic variation in *IRF5* could be. We also tested for a joint effect of *IRF5* and the STAT4 gene in one of the autoimmune diseases. Since *STAT4* had recently been shown to be associated with SLE, and along with *IRF5*, is involved in the type I interferon system this was of particular interest.

Aims

- To test polymorphisms in *IRF5* for association with inflammatory bowel diseases (Study I), multiple sclerosis (Study II) and primary Sjögren´s syndrome (Study IV)

- To identify all genetic variation of *IRF5*, to be able to identify the variants that are most strongly associated with SLE (Study III)

- To investigate joint effects of *IRF5* and *STAT4* in the disease susceptibility to primary Sjögen´s syndrome (study IV)

- To study the functions of the associated variants of *IRF5* (Study I-III)

Material and methods

DNA samples

In the four studies of the thesis, we analyzed a total of 8960 DNA samples: 4884 samples from patients and 4076 samples from controls. All study subjects provided informed consent to participation, and all studies were approved by the regional ethics boards.

IBD cohort

Two cohorts, in total 1694 IBD patients and 552 controls were included in this study. Both cohorts were recruited in Belgium. Our main cohort contained 1007 IBD patients from Wallonia, of which 748 were diagnosed with CD, 254 with UC and five with indeterminate colitis (IC). Healthy individuals attending the University Hospital of Liege and blood donors were used as controls (n=241). The second cohort consisted of 687 IBD patients from Leuven, of which 488 were CD patients, 192 were UC patients and 7 had IC. A set of 311 volunteer blood donors served as unrelated control individuals for these patients.

MS cohort

A total of 5150 MS patients and controls collected in Spain, Sweden and Finland were included in the study. All patients (n=2337), had clinically or laboratory-supported definite MS according to the Poser criteria[120] or fulfilled the criteria of McDonald for MS[121]. The Spanish cohort consisted of 660 MS patients and 833 blood donors as controls, recruited at three public hospitals in the south of Spain. The Swedish cohort consisted of 1166 MS patients and 1235 controls of Nordic ethnicity recruited at three hospitals in Stockholm. The Finnish cohort consisted of 511 MS trio families recruited at five hospitals in Finland.

SLE cohort

Our SLE study included 485 patients and 563 controls All patients fulfilled at least four of the classification criteria for having SLE as defined by the American college of Rheumatology (ACR)[82]. All patient samples were recruited from three University Hospitals in Sweden. The controls were recruited from healthy blood donors matched for sex, age and geographical region and from a population-based control cohort; Epidemiological Investigation of Rheumatoid Arthritis (EIRA)[122].

SS cohort

In the SS cohort a total of 368 patients with primary Sjögren's syndrome and 711 controls were included. All the patients fulfilled the American European

consensus criteria[88] for having primary Sjögren´s syndrome. The cohorts were collected in three university hospitals in Sweden (228 patients, 563 controls) and one university hospital in Norway (140 cases, 148 controls). The 563 Swedish controls included in this study were also included in study III.

Genotyping

The polymorphisms genotyped in Studies I, II and IV were selected because they had previously been shown to be associated with SLE, RA or had been suggested to influence the expression of *IRF5*. In Study III a comprehensive set of polymorphisms identified by sequencing on the IRF5 gene were genotyped.

Indels polymorphisms were genotyped either by size separation of the PCR products on high resolution agarose gel or by capillary electrophoresis on a DNA sequencing instrument.

SNPs were genotyped using three genotyping systems: GenomeLab™ SNPstream system from Beckman Coulter, florescent polarization template directed incorporation (FP-TDI) using reagents purchased from Perkin-Elmer or prepared "in house", or the GoldenGate assay and a BeadStation instrument from Illumina.

SNPstream and FP-TDI

Both the SNPstream and FP-TDI assays are enzymatic methods, where an extension primer is designed so that its 3´end hybridizes just adjacent to the SNP to be genotyped. Templates for the primer extension reaction are generated by standard PCRs. A DNA polymerase is then used to extend the primer over the SNP site, by incorporating fluorophore-labeled ddNTPs. By using labeled ddNTPs, further extension of the primer is inhibited, and the incorporated nucleotide becomes detectable. In the original publication this single-base primer extension method for genotyping is called "minisequencing"[123] (Figure 2).
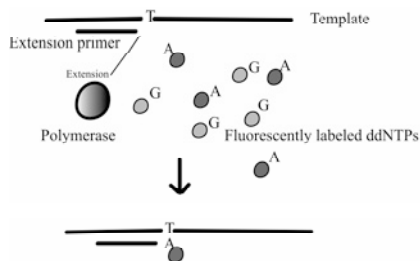

Figure 2. Principle of the minisequencing reaction. A minisequencing primer anneals adjacent to the SNP site to be genotyped. A DNA polymerase extends the primer by a single base, thus incorporating a fluorescently labeled ddNTPs.

Both SNPstream and FP-TDI use a dual color reaction, allowing both alleles of a SNP to be detected in the same reaction. The differences between the two methods are in the assay format and separation/detection step.
In FP-TDI, detection is achieved in solution by fluorescent polarization. This only allows one SNP to be genotyped at a time.

Using SNPstream, detection is performed by fluorescence on arrays, which allows multiplex genotyping of SNPs. The extension primers, used in the extension reaction, are constructed so that they contain tag-sequences that are complementary to tags-oligonucleotides immobilized on the arrays. In SNPstream from 12 to 48 SNPs can be processed in each well of an arrayed 384-well plate.

GoldenGate assay

The GoldenGate assay is based on allele-specific primer extension. In brief the procedure is as follows: Three different oligonuclotides are synthesized for each SNP; two allele-specific oligonucleotides primers that distinguish the SNP-alleles and a locus-specific oligonucleotides probe downstream of the SNP site. All the three oligonucleotides contain target sequences for a set of universal PCR primers. The locus specific oligonucleotide also contains specific address sequences that are complementary to oligonucleotide tags attached to beads. The oligonucleotides are allowed to hybridize directly to genomic DNA, and only the fully matched oligo is extended by the DNA polymerase, followed by ligation to the downstream probe. Each ligation product is then PCR-amplified with a fluorescently labeled universal primer pair. The PCR products are hybridized onto a bead array, and a BeadArray reader is used to measure the fluorescence on each bead (Figure 3). The maximal multiplexing level of the Golden Gate assay is 1536 SNPs in a 96 well format.

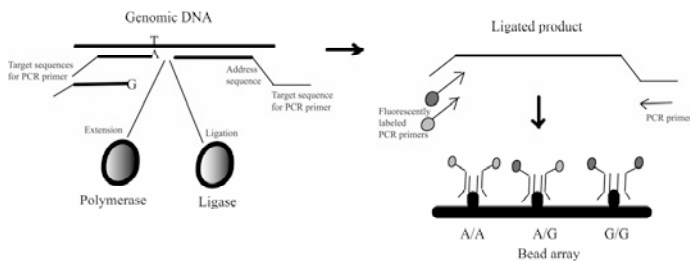**Figure 3**. Principle of the GoldenGate assay. Oligonucleotide primers and probes are hybridized to genomic DNA. Only an oligonucleotides that is prefectly matched with the SNP is extended and ligated to a locus-specific probe. The ligation product is amplified with flurorescent universal primers. The PCR products are guided to the beads with the aid of the address sequence on the locus specific oligonucleotides probe.

Sequencing

In Study III we sequenced the IRF5 gene in eight controls and 40 SLE cases. The whole gene was sequenced, using 23 PCR fragments that covered 1kb upstream of exon 1a and all exons and introns. The primers for PCR were designed using the Primer3 software. PCR was performed and the PCR products were cleaned with Exonuclease I and Shrimp Alkaline Phosphatase mix and sequenced at the sequencing facility at the Rudbeck laboratory in Uppsala. The sequence reads were analyzed using the Sequencher version 4.5. software.

Electrophoretic mobility shift assay (EMSA)

We performed EMSA in each of the three first studies to investigate if polymorphisms in *IRF5* found to be associated with the diseases, affect protein binding. For each of the polymorphisms studied, pairs of single stranded 5´-biotin-labelled and unlabeled oligonucleotides were allowed to anneal to generate double-stranded probes. Labeled probes were incubated with nuclear extract prepared from PBMCs. To estimate preference of binding we also performed competition experiments, where we added a 100-fold molar excess of unlabeled probe to the incubation. The DNA-protein complexes produced in the binding reactions were analyzed using electrophoresis on polyacrylamide gels. After the electrophoretic separation, the biotinylated fragments were transferred to membranes, and detected by a chemiluminescent procedure using the LightShift® Chemiluminescent EMSA kit.

Proximity ligation assay (PLA)

In Study II we used PLA[124] to confirm our hypothesis that the SP1 protein binds to the polymorphic CGGGG repeat. The power of this technique is based on two bi-functional probes. One of the probes is a labeled antibody directed against the protein of interest that has been conjugated to an oligonucleotides. Here we used biotinylated polyclonal antibody against the SP1 protein, combined with a streptavidin–oligonucleotide conjugate. The other probe consisted of a partially double-stranded DNA sequence containing the SP1 binding site. Here we used a DNA probe containing the polymorphic CGGGG repeat purified by high pressure liquid chromatography (HPLC). The HPLC-purified probe was made partially double stranded as described by Gustafsdottir et al [124]. Since the SP1 protein was simultaneously bound by both probes, the oligonucleotides ends of the probes were brought physically close together and were thus able to hybridize to added to a connector oligonucleotide. This DNA structure was then covalently joined by enzymatic ligation. The ligated DNA sequence, which serves as a representation of the binding event between SP1 and the CGGGG repeat, was then amplified and detected by real-time PCR.

Figure 4. Analysis of SP1 protein interaction by proximity ligation. Two affinity probes, an antibody against the SP1 protein and a DNA oligo that contain either allele of the indel, are bound simultaneously to the SP1 protein. Both probes carry an oligonucleotide extension. The vicinity of the probes leads to hybridization of their oligonucleotide extension to added connector oligonucleotides, allowing them to join by an enzymatic reaction. The ligated DNA sequence is amplified and detected by real time PCR.

Minigen reporter assays

In Study III we compared the two alleles of the 5bp indel polymorphism with regards to expression of the IRF5 gene, in SLE patients. DNA fragment containing 1.4kb upstream of exon 1a and part of exon 2 of *IRF5* was generated by PCR from genomic DNA. The fragments were cloned into a pCR-Blunt II TOPO vector. After verification by sequencing, the fragments were subcloned into a pcDNA3 vector and transfected into cells lacking endogenous *IRF5* expression. Total RNA was isolated from the transfected cells and cDNA corresponding to exon 1a-associated transcripts, was synthesized from the total RNA. To compare the amount of *IRF5* expression from promoters containing either 3 or 4 repeats of the 5bp CGGGG indel, the cDNAs were subjected to real-time PCR. The specificity of each reaction was examined by dissociation curve analysis using the pcDNA3 vector or β-actin for standardization.

Immunoblot analysis

Immunoblot analysis was performed to confirm the results obtained using the minigen reporter assay, namely increased expression from *IRF5* containing the risk allele of the 5bp CGGGG indel polymorphism. The principle of this technique is, first to detect a specific protein by means of its molecular mass and then by its antigenicity. The immunoblot analysis was performed in the following way: Cell lysates from purified peripheral blood mononuclear cells pellets were prepared and denatured. Proteins were separated by Sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (PAGE). After separation, the proteins were transferred onto a nitrocellulose membrane and probed with

mouse anti-human IRF5 antibodies. To be confident that each sample had been loaded with equal amounts of cell extract, the membrane was also probed with rabbit anti-alpha-tubulin as an internal control. Horseradish peroxidase-conjugated anti-mouse and anti-rabbit secondary antibodies were used to detect the protein-primary antibody complexes formed. Immunoreactive protein complexes were visualized with enhanced chemiluminescence using ECL reagents.

Statistical analysis

A χ^2 test (P>0.05) was used to assess that the genotype distributions of the polymorphisms fulfilled the criteria of Hardy-Weinberg equilibrium. The PLINK software was used to compare the allele counts in cases and controls by Fisher´s exact test and to calculate odds ratios (OR). Plink was also used to perform haplotype association analysis in Study I and II and conditional logistic regression analysis in Study III. The Haploview software was used to determine linkage disequilibrium (LD) between the polymorphisms. A Bayesian model selection approach[125] was used to identify genetic variants that have the highest joint probability of being associated with SLE. For the family cohort in Study II, the genotype data was analyzed using the TRANSMIT 2.5.2. software. Alleles that were included in the analysis were required to be transmitted from a least 20 informative meioses. A combined p-value from case-control cohorts and trios were calculated using the analytical expression by Jost, $k(1-(\ln(k))+(-\ln(k)^2)/2)$, where k is the product of the p-values from the different cohorts (http://www.loujost.com). The ssSNPer tool was used to determine the pair-wise correlation between SNPs in the HapMap data[126].

Results

Association analysis of *IRF5*

<u>Study I: IBD</u>

In Study I we tested 12 polymorphisms, ten SNPs and two insertion-deletions, in the IRF5 gene (Figure 5) for their association with two subgroup of inflammatory bowel diseases; Crohn´s disease and ulcerative colitis. The polymorphisms were selected because they had previously been shown to be associated with SLE, RA or had been suggested to modulate the expression of *IRF5*.
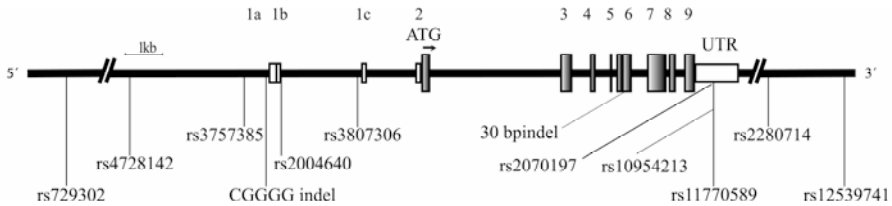

Figure 5. Schematic illustration of the IRF5 gene with the positions of the analyzed polymorphisms. The exons and 3´-UTR are shown as boxes. The exons are labeled as 1-9 and the untranslated alternative exons are labeled as 1a, 1b and 1c. The transcription initiation site is indicated with ATG and an arrow above the gene.

We found six out of the 12 polymorphisms to be associated with IBD (p< 0.05), in our initial cohort of 987 cases and 223 controls. All six polymorphisms are located in the promoter region and the first intron of the gene.
The same six polymorphisms showed significant association signals in both subgroups of IBD, CD and UC. The strongest signal was obtained for the CGGGG indel and this signal was particularly strong in the UC group with a p-value of 5.3×10^{-8} and an OR of 2.4.

To confirm the association of *IRF5* to IBD we used an independent cohort of 687 IBD patients and 311 controls. In the confirmational cohort the CGGGG indel was found to be strongly associated with IBD (p-value 3.2×10^{-5}, OR 1.59), while no association signal was observed for the other polymorphisms.

When we combined the two cohorts, strong association signals from the CGGGG indel polymorphism and from the SNP rs4728142 were further strengthened (Table 1). Both patient subgroups were also considered in the combined analysis. As can be seen in Table 1, a particularly strong signal of association was observed for both polymorphisms in the UC group.

Table 1. Single marker association results from combined cohort of 1661 inflammatory bowel disease (IBD) patients and 534 controls. Of the patients 1207 were diagnosed with Crohn´s disease (CD) and 429 with ulcerative colitis (UC).

Polymorphism	Risk allele	Allele frequency				P-value			Odds ratio (95%CI)		
		Controls	IBD	CD	UC	IBD	CD	UC	IBD	CD	UC
rs729302	A	0.38	0.36	0.37	0.32	2.60E-01	8.10E-01	1.10E-02	0.91(0.79-1.07)	0.98(0.83-1.15)	0.77(0.63-0.94)
rs4728142	A	0.38	0.42	0.42	0.48	1.40E-03	2.90E-02	4.20E-05	1.28(1.1-1.49)	1.20(1.02-1.40)	1.50(1.24-1.83)
rs3757385	G	0.4	0.37	0.37	0.37	1.90E-01	3.00E-01	2.80E-01	0.90(0.78-1.05)	0.92(0.78-1.07)	0.90(0.74-1.09)
5bp indel	Ins	0.67	0.56	0.43	0.49	1.40E-08	3.30E-06	7.90E-10	1.62(1.37-1.91)	1.51(1.27-1.80)	1.93(1.56-2.38)
rs2004640	T	0.47	0.5	0.49	0.53	1.60E-01	4.60E-01	2.70E-02	1.11(0.96-1.29)	1.06(0.91-1.24)	1.24(1.03-1.50)
rs3807306	T	0.44	0.48	0.47	0.52	2.70E-02	1.70E-01	1.80E-03	1.18(1.02-1.37)	1.11(0.95-1.30)	1.36(1.12-1.64)
30bp indel	Del	0.5	0.48	0.49	0.47	3.00E-01	5.10E-01	2.70E-01	0.92(0.80-1.07)	0.95(0.81-1.10)	0.90(0.74-1.08)
rs2070197	C	0.081	0.094	0.094	0.094	2.70E-01	2.70E-01	3.90E-01	1.17(0.89-1.52)	1.17(0.89-1.54)	1.17(0.84-1.67)
rs10954213	A	0.42	0.39	0.39	0.38	1.10E-01	2.10E-01	1.50E-01	0.88(0.76-1.03)	0.90(0.77-1.06)	0.87(0.71-1.05)
rs11770589	A	0.5	0.49	0.49	0.48	3.00E-01	4.80E-01	2.90E-01	0.92(0.80-1.07)	0.95(0.81-1.10)	0.90(0.75-1.09)
rs2280714	T	0.36	0.35	0.35	0.35	3.10E-01	4.10E-01	4.80E-01	0.92(0.79-1.08)	0.93(0.80-1.10)	0.93(0.77-1.13)
rs12539741	T	0.077	0.095	0.096	0.095	1.10E-01	1.10E-01	1.90E-01	1.25(0.95-1.65)	1.26(0.95-1.68)	1.26(0.90-1.78)

We performed a haplotype association test of the six polymorphisms that had shown association with IBD in our initial cohort, to investigate whether the haplotype could capture the association signal at a higher significance than the individual polymorphisms. This analysis did not reveal any association signals that were stronger than those for the CGGGG indel alone.

To assess to which extent the associated allele of the CGGGG indel confers risk to IBD, we calculated the OR for having a single or two copies of the risk allele. This resulted in OR 1.62 (1.37-1.91) for having one copy and OR of 3.03 (2.04-4.49) if having two copies of the risk allele.

Study II: MS

Ten out of the 12 polymorphisms from Study I were genotyped in the MS cohorts. One SNP (rs2070197) and a 30bp insertion deletion were omitted. The reason for leaving out those two polymorphisms was that the SNP is in perfect LD with the SNP rs12539741, which was included in the study, and that the 30bp indel did not reveal any association with IBD in Study I.

The ten selected polymorphisms were first genotyped in a Spanish cohort of MS patients (n=660) and controls (n=833). Seven of the ten polymorphisms showed significant signals of association with MS (p<0.05).To replicate our findings, we genotyped the same set of polymorphisms in an independent case-control cohort of 1166 MS patients and 1235 matched controls from Sweden. Two of the SNPs (rs4728142 and rs3807306), and the 5 bp CGGGG indel polymorphism, showed a nominally significant association (p<0.05). Each of these polymorphisms where also associated in the Spanish cohort.

In a further attempt to verify our findings, we genotyped 511 Finnish MS trio families. Using a transmission disequilibrium test, 4 SNPs showed association with MS in the Finnish cohort (p<0.05). The 5 bp indel did not reach the limit for association but its p-value was 0.056. It should be noted however, that the overtransmitted allele in the trio analysis is the same as the risk allele in the case-control cohorts.

When the data from all cohorts are combined, association signals from the two SNPs (rs4728142 and rs3807306) and the CGGGG indel were maintained (Table 2).

Table 2. Combined association analysis of *IRF5* polymorphisms with multiple sclerosis in three populations.

Polymorphism	Spanish c/c[1] P-value	Swedish c/c[1] P-value	Finnish trios P-value	Combined P-value
rs729302	0.22	0.29	0.047	0.07
rs4728142	**0.0029**	**0.02**	**0.035**	**0.0002**
rs3757385	0.0052	0.43	0.53	0.04
5bp indel	**0.011**	**0.0093**	**0.056**	**0.0005**
rs2004640	0.0011	0.082	0.54	0.003
rs3807306	**0.0037**	**0.049**	**0.012**	**0.0002**
rs10954213	0.031	0.15	0.01	0.003
rs11770589	0.29	0.49	0.07	0.16
rs2280714	0.003	0.43	0.3	0.02
rs12539741	0.074	0.1	0.74	0.11

[1]c/c denotes case-controls.

The three polymorphisms, rs4728142, rs3807306 and the 5bp indel, are in high LD with each other, with r^2 values of 0.61-0.88. A haplotype analysis revealed that the risk alleles of the polymorphisms were all present on the same common haplotype in the Spanish, Swedish and Finnish populations.

Study III: SLE
Despite the well established association of the IRF5 gene with SLE, it is not apparent which the true associated functional polymorphism in the gene is. In Study III we therefore aimed to identify all the genetic variation in *IRF5* and clarify which genetic variants are jointly most strongly associated with SLE.

The whole IRF5 gene was sequenced in 48 individuals; 40 SLE patients and 8 controls. Forty eight polymorphic sites were identified. No missense SNPs in the protein coding region of *IRF5* was discovered. Based on LD and GoldenGate assay score, 27 of the polymorphisms were selected for genotyping. In addition, 20 SNPs from dbSNP located within 9kb upstream or 5kb downstream from the gene were included in our genotyping panel. In this panel a total of 47 polymorphisms were genotyped in 485 SLE patients and 563 controls, and tested for their association with SLE.

Eighteen of the polymorphisms yielded association signal with SLE, even after Bonferroni correction for multiple testing. A particularly strong association signal was observed for the 5bp CGGGG indel polymorphism located upstream of exon 1a (p= 4.6×10^{-9} and OR = 1.69) and the SNP rs10488631 located 5kb downstream of the gene.

By using conditionally logistic regression analysis it was observed that the 5bp indel and SNP rs10488631 together eradicate the association signal from all the remaining markers in the IRF5 gene, and that these two polymorphisms are independently associated with SLE (Table 3). Since only a subset of all possible combinations can be examined using conditional logistic regression, a Bayesian model selection analysis was also performed on the data. With this computer-intensive method all possible combinations of polymorphism can be analyzed, since the model does not require previous selection of a conditioning variant. Both methods gave similar results.

Table 3. *IRF5* polymorphisms associated with SLE with P<0.0005 and conditional logistic regression analysis.

| | | P values for conditional logistic regression [b] | | |
| | Single marker | | | CGGGG indel and |
Marker	P-values [a]	CGGGG indel	rs10488631	rs10488631
rs729302	2.7E-04	0.22	0.04	0.63
rs4728142	1.8E-07	0.31	6.20E-04	0.63
rs3778754	1.8E-07	0.28	2.90E-03	0.89
5bp indel	4.6E-09	NA	5.80E-05	NA
rs2004640	5.7E-07	0.23	3.00E-03	0.83
rs752637	1.8E-04	0.47	0.03	0.94
rs3807306	3.9E-07	0.24	3.10E-03	0.84
rs11761199	4.9E-06	0.58	0.02	0.51
rs7808907	4.4E-06	0.25	0.01	0.96
Gindel	8.7E-06	5.70E-03	0.61	0.92
rs10954213	8.6E-05	0.41	0.03	0.89
rs10954214	1.5E-04	0.3	0.03	0.69
rs13242262	1.2E-04	0.45	0.04	0.94
rs7800687	1.8E-04	0.32	0.03	0.71
rs10488631	9.4E-10	1.90E-05	NA	NA
rs2280714	2.1E-04	0.35	0.04	0.76

[a] P values based on comparison of allele count between cases and controls by Fisher´s exact test.
[b] P values for association conditional on the indicated marker.

Study IV: pSS

We investigated the association of the two polymorphisms in the IRF5 gene and one polymorphism in *STAT4*, with primary Sjögren´s syndrome in a cohort of 368 well characterized pSS patients and 711 controls. The genotyped polymorphisms in the IRF5 gene were selected based on our result from Study III, were we found that the 5bp indel and one SNP rs10488631, are independently associated and account for most of the observed association signals of *IRF5* with SLE. The single polymorphism selected in the STAT4 gene (rs7582694) has previously been shown to account for the entire association signal with SLE[127].

The result from our study was that all three polymorphisms showed strong signals for association with Sjögren´s syndrome in the Swedish cohort alone. In the Norwegian cohort, the 5bp indel was significantly associated

with pSS and the *IRF5* SNP rs10488631 showed a suggestive association with the disease. The reason for the low association signals of *IRF5* with pSS in the Norway could be the low number of individuals in that cohort or true differences between the populations. In meta-analysis of the data from both cohorts, the significance of the association with primary SS was strengthened for all three polymorphisms. The association signals in the combined cohort were slightly stronger for the two polymorphisms in *IRF5* with an OR of 1.49 for the 5bp indel and an OR of 1.57 for the SNP rs10488631, compared to an OR of 1.41 for the SNP rs7582694 in *STAT4* (Table 4). It is notable that there was a slight difference in the risk allele frequencies for the *IRF5* polymorphisms between individuals from Norway and Sweden. We observed higher risk allele frequencies in Norway in both primary SS patients and controls, while there was no difference in risk allele frequencies between the two populations for the *STAT4* SNP.

Table 4. Association analysis of three polymorphisms with primary Sjögren´s syndrome in Swedish and Norwegian cohorts.

| | | Sweden | | | | Norway | | | | Combined cohorts | |
| | | Risk allele frequency | | | | Risk allele frequency | | | | | |
Marker	Risk allele	Cases (n=228)	Controls (n=563)	P-value	OR	Cases (n=140)	Controls (n=148)	P-value	OR	P-value	OR 95% CI
5bp indel	4x	0.54	0.44	2.0×10^{-4}	1.51	0.59	0.5	0.03	1.45	2.41×10^{-5}	1.49 (1.24-1.79)
rs10488631	C	0.19	0.12	1.4×10^{-3}	1.62	0.25	0.18	0.06	1.47	3.23×10^{-4}	1.57 (1.23-1.99)
rs7582694	C	0.29	0.22	4.7×10^{-3}	1.43	0.29	0.23	0.11	1.36	1.45×10^{-3}	1.41 (1.14-1.73)

Joint effect of *IRF5* and *STAT4* in Sjögren´s syndrome

We also investigated the joint effect of the two polymorphisms in *IRF5* since we assumed that they give rise to independent association signals with pSS. We found that the risk haplotype carrying both risk alleles had a p-value of 9.57 x 10^{-5} and an OR of 1.63. The protective haplotype on the other hand, carrying the non-risk alleles had a p-value of 2.88 x 10^{-5} and an OR of 0.67. The haplotypes formed by one risk allele and one non-risk allele of *IRF5* was neutral (Table 5). These results are consistent with the independent effect on primary SS by these two polymorphisms.

Table 5. Association analysis of *IRF5* haplotypes with primary Sjögren´s syndrome in the Swedish and Norwegian cohorts.

| | Sweden | | | | Norway | | | | Combined cohort | |
| | Haplotype frequency | | | | Haplotype frequency | | | | | |
Haplotype	Cases	Controls	P-value	OR	Cases	Controls	P-value	OR	P-value	OR 95% CI
3xindel-T	0.45	0.55	3.00×10^{-4}	0.67	0.39	0.48	0.03	0.68	2.88×10^{-5}	0.67 (0.56-0.81)
4xindel-T	0.37	0.33	0.16	1.18	0.36	0.34	0.52	1.12	1.35×10^{-1}	1.16 (0.96-1.40)
4xindel-C	0.18	0.11	5.00×10^{-4}	1.71	0.23	0.17	0.05	1.53	9.57×10^{-5}	1.63 (1.28-2.09)
3xindel-C	0.01	0.01	0.52	0.7	0.02	0.02	0.93	1.07	8.12×10^{-1}	0.82 (0.35-1.94)

We also tested for a joint effect of the risk alleles of both genes, *IRF5* and *STAT4* on the disease susceptibility to pSS. The distribution of pSS patients and controls in each group of individuals carrying the different numbers of risk alleles are plotted in figure 6. The overall differences in risk allele count between patients and controls are very high, with a one sided P-value of 2.5×10^{-9}.
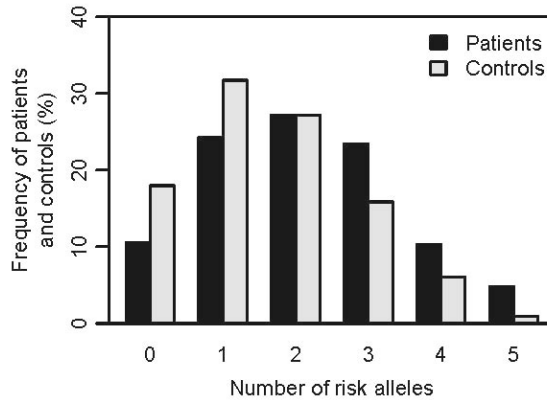

Figure 6. Frequency of pSS patients and controls carrying different numbers of risk alleles. Black bars indicates patients and the gray bars indicates controls. No individual carried all six risk alleles.

In Figure 7 we visualize the relationship between the number of carried risk alleles and the risk for developing pSS. The resulting straight line shows that the risk for developing the disease increases in an additive manner, with an average increase in OR of 1.78 for each risk allele. For carriers of two risk alleles the OR for primary SS is 1.43, while carriers of 5 risk alleles have an OR of 6.78.
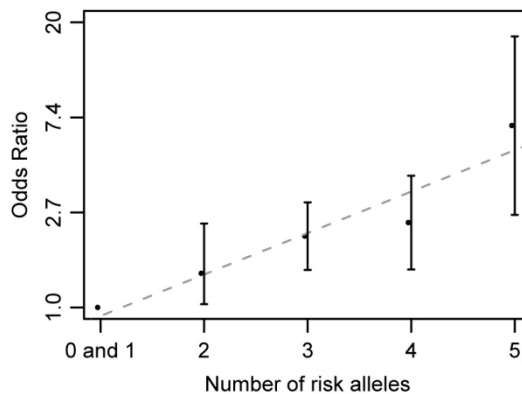
Figure 7. Linear regression analysis showing the odds ratios (ORs) with 95% confidence intervals for developing pSS as a function of the number of risk alleles. The scale of the y-axis corresponds to the natural logarithm of the ORs.

We tested for dominance or interaction between the risk alleles by using multiple logistic regression analysis. No evidence of such interactions was observed.

It has been shown that sera from pSS patients in combination with necrotic or apoptotic cell material induce IFN-α production in peripheral blood mononuclear cells (PBMCs) *in vitro* and this IFN-inducing ability is correlated with the presence of anti-SSA and -B antibodies[113]. Inspired by this knowledge, we investigated if there is a correlation between the presence of these antibodies and the three polymorphisms under investigation. No correlation was found between the presence of ANA, anti-SSA or anti-SSB antibodies and the genotypes of the *IRF5* or *STAT4* polymorphisms.

Functional analysis of *IRF5*

We were interested to determine a possible functional role for the 5bp indel in the IRF5 gene, since this variant has been shown to be associated in all the diseases studied in this thesis. We knew from sequencing of *IRF5,* that the indel contains either 3 or 4 CGGGG repeats and from the association analysis we knew that the insertion of one repeat (4xCGGGG) is the risk allele. We used electrophoretic mobility shift assays (EMSA) to compare protein binding to the alleles of the indel. This analysis revealed a higher level of binding of protein to the risk allele than to the non-risk allele (Figure 8).

3x CGGGG   4x CGGGG
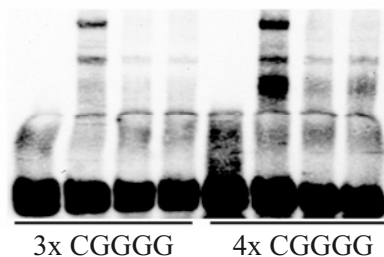
Figure 8. EMSA images for the 5bp indel polymorphisms. Samples loaded in lane 1-4 contains for each allele: (1) labeled probe, (2) labeled probe and nuclear extract, (3) labeled probe, nuclear extract and unlabeled probe, (4) labeled probe, nuclear extract and 100 fold excess of unlabeled probe for the other allele of the polymorphisms added as cross competitor.

A limitation of the EMSA method is that no information regarding the identity of the protein that binds to the DNA is obtained. By searching a transcription factor database, we found that the sequence of the CGGGG indel is predicted to bind the transcription factor SP1. The risk allele of the 5bp polymorphisms with 4 CGGGG repeats contains three SP1 binding sites, while the non-risk allele with 3 CGGGG repeats is predicted to contain two such sites. To verify that it is the SP1 protein that actually binds to the indel po-

lymorphism, we used an antibody against SP1 in the proximity ligation assay. The results showed a 2.6 fold higher binding signal from the 4x CGGGG probe (the risk allele) than from the probe containing the 3x CGGGG sequence (non risk allele), confirming that the SP1 protein binds to the indel polymorphisms (Figure 9).
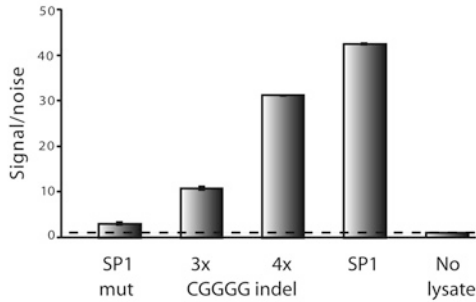

Figure 9. Analysis of the SP1 protein-promoter interaction using the proximity ligation assay.

We also investigated if the allele-specific difference in protein binding, which we had observed in our EMSA experiment, had an effect on the expression level of *IRF5*. Two experiments were conducted to test this, minigene reporter assays and immunoblot analysis.
In both of these experiments it was observed that the risk allele of the CGGGG indel is associated with increased expression of the IRF5 gene. In the minigene assay, promoters from individuals homozygous for either allele of the indel were cloned into expression vectors. *IRF5* expression was determined to be two to tenfold higher for promoters carrying the risk allele.

Summary of the results

Significant association with interferon regulator factor 5 was found in all the diseases analyzed, inflammatory bowel disease, multiple sclerosis, systemic lupus erythematosus and primary Sjögren´s syndrome.

- **Study I** provides the first confirmed evidence of a shared gene between SLE and IBD. We found *IRF5* to be associated with both CD and UC. To our knowledge this is the first gene associated with IBD that shows a stronger association signal with UC than with CD.

- **Study II** shows association of *IRF5* with MS in three different European populations; the Spanish, Finnish and Swedish populations. This finding adds *IRF5* to the short list of genes shown to be associated with MS in more than one population.

- In **Study III** we identified a complete set of polymorphisms in the IRF5 gene, 30 SNPs and 4 indels, and showed association with SLE for 18 of those polymorphisms. Two polymorphisms (SNP rs10488631 and a 5bp CGGGG indel) in the gene were found to contribute to all the association signals obtained with SLE.

- In **Study IV** we show an association of *IRF5* with primary Sjögren´s syndrome in two Scandinavian populations, the Swedish and the Norwegian populations. We also showed association of the STAT4 gene with pSS, and demonstrated a strong additive affect of the risk alleles of the two genes in primary Sjögren´s syndrome.

In Studies I, II and III we performed functional analysis of the most strongly associated polymorphism in the IRF5 gene, the 5bp CGGGG indel. The results showed increased expressions of *IRF5*, as a consequence of increased binding of SP1 protein to the risk allele of the indel polymorphism.

Discussion

By analyzing a comprehensive set of polymorphisms in the IRF5 gene, we found that two main groups of polymorphisms account for the obtained association signal with SLE. The former group is defined by a 5bp CGGGG indel in the promoter region of the gene, and the second group is defined by a SNP rs10488631 located 5kb downstream of the gene. The SNP rs10488631 is a perfect proxy for several polymorphisms in the 3´end of *IRF5* as well as SNPs distributed over a 100 kb region downstream of the gene, where the transportin 3 (TNPO3) gene is located. By genotyping methods it is impossible to clarify if this second signals originates from the IRF5 gene, or if it actually originates from polymorphisms in *TNPO3,* or from other polymorphisms located downstream of *IRF5*.

Interestingly in IBD and MS, we observe association signals from the 5bp indel, but not from the two SNPs rs2070197 and rs12539741 which are according to HapMap in almost perfect LD, with SNP rs10488631. Our data shows clearly that the association signal that we obtain from *IRF5* with these two diseases, only originates from polymorphisms belonging to the former group, which clarifies that it is *IRF5*, and not *TNPO3* that is associated with IBD and MS in our study.

In our investigation of primary SS, we found association with both the 5bp indel and the SNP rs10488631. These results are highly similar to those found in SLE. There is however, a publication showing association of IRF5 with primary Sjögren´s syndrome only for the group of polymorphisms that is tagged by the 5bp indel, and no association with the second group of SNPs[58]. The discrepant results can possibly be explained by the small size of that study, and thereby too little power to detect the signal from the second group.

In our study on primary SS we also analyzed the association of *STAT4* with the disease and found the same risk allele that has been shown to be associated with both SLE and RA to be associated also with pSS. This implies that both *IRF5* and *STAT4* and the type I IFN system are important in the development of rheumatic diseases.

We did not investigate association of *STAT4* with IBD or MS, which should be of great interest. A recent publication on the association of *STAT4* with several immune mediated disease reported an association of *STAT4* with IBD but did not detect an association with MS[128].

Using statistical methods in Study III, we found that two groups of polymorphisms contribute independently to the association signals observed

from the IRF5 gene. As previously listed, these groups are represented by the 5bp indel polymorphisms and the SNP rs10488631 respectively. Our results are based on the association data, and do not give evidence of a disease causing role for these polymorphisms. The best way to identify the real causal polymorphisms is by functional studies. Several polymorphisms in *IRF5* belonging to both of these groups have been identified to have functional roles of the gene.

In the former group, the 5bp indel polymorphism is the most likely causal variant. The indel is located in the promoter region of the gene, only 64 bp upstream of exon 1a. Our functional analyses support a role for the 5bp indel as a *cis*-acting regulatory element, based on the prediction that the risk allele of the 5bp indel creates an additional binding site for the transcription factor SP1. This sequence-based prediction was verified experimentally by showing increased SP1 protein binding to the risk allele compared to the non-risk allele.

The SP1 protein is known to interact with the IFN stimulated response element (ISRE), which is found in the promoter region of IFN inducible genes, and that increases the expression of at least some of these genes[129]. This could also be the case for IRF5, which, as other IRFs, contains an ISRE motif in the promoter region. The result from the minigene analyses support the role of the 5bp indel as regulator of *IRF5* expression.

Our results do not exclude a functional role for other disease-associated polymorphisms in *IRF5*. Other possible functional polymorphisms belonging to the same group as the 5bp indel, are the SNP rs2004640, and the SNP rs10954213. The SNP rs2004640 is located at the exon-intron splice junction of the alternative exon 1b of *IRF5* and has been shown to be associated with the transcription level of *IRF5* mRNAs containing alternative exon 1b[130]. The SNP rs10954213 in the 3'-UTR of IRF5 has been suggested to correlate with the stability of IRF5 mRNA in the cell[131]. It should be noted that at least nine splice variant have been detected from the IRF5 gene, with alternative three first exons[119]. In peripheral blood mononuclear cells the most abundant *IRF5* transcript contains exon 1a, and the expression level of this transcript has been reported to be independent of the genotype of SNP rs2004640[132]. This finding speaks against rs2004640 as major functional variant of *IRF5* expression. It is possible however, that alternative spliced transcript containing exon 1b is expressed at higher levels in other tissues.

Second group of possible functional variants in the IRF5 gene are known and unknown proxies to the SNP rs10488631. No obvious functional roles for these variants have been predicted.

Further functional studies are needed to clarify the role of *IRF5* in the pathogenesis of autoimmune diseases. Several pathogenic mechanisms are possible. For example, increased expression of *IRF5* could lead to increased production of pro-inflammatory cytokines and thereby maintenance of inflammation and enhanced apoptosis, or a differential immune response depending on the expression of *IRF5* isoforms.

To identify polymorphisms that have functional consequences and a putative role in autoimmunity is of great importance. Such future studies could be paving the way for discovery of novel immunomodulatory treatments or even choosing suitable pharmacological therapy for patients.

Acknowledgements

Anders, for assisting with the statistics and Raul for all your help with the blood collecting and DNA/RNA isolations. Lillebil for taking care of the administrative work in our group and Lars for helps solving computer problems.

Vinciane Dideberg, thank you for the wonderful collaboration work on the IBD paper and the good time in Spain.

My room members Mohammed, Mina, Magnus and Anna (again!). For making our office such a pleasant working environment.

Jóhanna, Stína, Gunnhildur-Halla, and all my other Icelandic friends in Uppsala. Thank you for all the nice times we have spent together during our stay in Sweden. And the ones that have moved home; Sigga, Maja and Gunnar Holmsteinn, your are really missed in our Swedish/Icelandic community.

My family, Össi, Gudrun, Þorri and Krissi, without you I would not have moved to Uppsala. And without your help and understanding I could not have done this. Takk Guðrún Sara mín fyrir öll þau skipti sem þú hefur náð í bræður þína í skólann og hugsað svo vel um þá. Þetta hefði verið mun erfiðara ef þú hefðir ekki verið svona dugleg að hjálpa til.

References

1. Avery OT, MacLeod CM, McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus type III. J Exp Med 79:137-159
2. Watson JD, Crick FH (1953) Genetical implications of the structure of deoxyribonucleic acid. Nature 171:964-967
3. Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171:737-738
4. (2004) Finishing the euchromatic sequence of the human genome. Nature 431:931-945
5. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, et al. (2007) The diploid genome sequence of an individual human. PLoS Biol 5:e254
6. Ng PC, Levy S, Huang J, Stockwell TB, Walenz BP, Li K, Axelrod N, Busam DA, Strausberg RL, Venter JC (2008) Genetic variation in an individual human exome. PLoS Genet 4:e1000160
7. Bhangale TR, Rieder MJ, Livingston RJ, Nickerson DA (2005) Comprehensive identification and characterization of diallelic insertion-deletion polymorphisms in 330 human candidate genes. Hum Mol Genet 14:59-69
8. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 16:1182-1190
9. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. Science 305:525-528
10. Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA (2006) Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat Genet 38:82-85
11. Altshuler D, Daly MJ, Lander ES (2008) Genetic mapping in human disease. Science 322:881-888
12. Trofatter JA, MacCollin MM, Rutter JL, Murrell JR, Duyao MP, Parry DM, Eldridge R, Kley N, Menon AG, Pulaski K, et al. (1993) A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. Cell 75:826
13. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073-1080
14. Purrello M, Nussbaum R, Rinaldi A, Filippi G, Traccis S, Latte B, Siniscalco M (1984) Old and new genetics help ordering loci at the telomere of the human X-chromosome long arm. Hum Genet 65:295-299
15. Devys D, Lutz Y, Rouyer N, Bellocq JP, Mandel JL (1993) The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. Nat Genet 4:335-340

16. Howell N, Bindoff LA, McCullough DA, Kubacka I, Poulton J, Mackey D, Taylor L, Turnbull DM (1991) Leber hereditary optic neuropathy: identification of the same mitochondrial ND1 mutation in six pedigrees. Am J Hum Genet 49:939-950

17. Fossdal R, Jonasson F, Kristjansdottir GT, Kong A, Stefansson H, Gosh S, Gulcher JR, Stefansson K (2004) A novel TEAD1 mutation is the causative allele in Sveinsson's chorioretinal atrophy (helicoid peripapillary chorioretinal degeneration). Hum Mol Genet 13:975-981

18. (2003) The International HapMap Project. Nature 426:789-796

19. Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516-1517

20. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308:385-389

21. Donnelly P (2008) Progress and challenges in genome-wide association studies in humans. Nature 456:728-731

22. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, Vermeire S, Dewit O, de Vos M, Dixon A, Demarche B, Gut I, Heath S, Foglio M, Liang L, Laukens D, Mni M, Zelenika D, Van Gossum A, Rutgeerts P, Belaiche J, Lathrop M, Georges M (2007) Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. PLoS Genet 3:e58

23. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW, Shugart YY, Griffiths AM, Targan SR, Ippoliti AF, Bernard EJ, Mei L, Nicolae DL, Regueiro M, Schumm LP, Steinhart AH, Rotter JI, Duerr RH, Cho JH, Daly MJ, Brant SR (2007) Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet 39:596-604

24. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, et al. (2007) Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet 39:830-832

25. Raelson JV, Little RD, Ruether A, Fournier H, Paquin B, Van Eerdewegh P, Bradley WE, Croteau P, Nguyen-Huu Q, Segal J, Debrus S, Allard R, Rosenstiel P, Franke A, Jacobs G, Nikolaus S, Vidal JM, Szego P, Laplante N, Clark HF, Paulussen RJ, Hooper JW, Keith TP, Belouchi A, Schreiber S (2007) Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. Proc Natl Acad Sci U S A 104:14747-14752

26. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, et al. (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 40:955-962

27. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661-678

28. Franke A, Balschun T, Karlsen TH, Sventoraityte J, Nikolaus S, Mayr G, Domingues FS, Albrecht M, Nothnagel M, Ellinghaus D, Sina C, Onnie CM, Weersma RK, Stokkers PC, Wijmenga C, Gazouli M, Strachan D, McArdle WL, Vermeire S, Rutgeerts P, Rosenstiel P, Krawczak M, Vatn MH, Mathew CG, Schreiber S (2008) Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. Nat Genet 40:1319-1323

29. Silverberg MS, Cho JH, Rioux JD, McGovern DP, Wu J, Annese V, Achkar JP, et al. (2009) Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. Nat Genet 41:216-220

30. Franke A, Hampe J, Rosenstiel P, Becker C, Wagner F, Hasler R, Little RD, Huse K, Ruether A, Balschun T, Wittig M, Elsharawy A, Mayr G, Albrecht M, Prescott NJ, Onnie CM, Fournier H, Keith T, Radelof U, Platzer M, Mathew CG, Stoll M, Krawczak M, Nurnberg P, Schreiber S (2007) Systematic association mapping identifies NELL1 as a novel IBD disease gene. PLoS ONE 2:e691

31. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barmada MM, Rotter JI, Nicolae DL, Cho JH (2006) A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 314:1461-1463

32. Kugathasan S, Baldassano RN, Bradfield JP, Sleiman PM, Imielinski M, Guthery SL, Cucchiara S, et al. (2008) Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. Nat Genet 40:1211-1215

33. Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, de Bakker PI, Gabriel SB, Mirel DB, Ivinson AJ, Pericak-Vance MA, Gregory SG, Rioux JD, McCauley JL, Haines JL, Barcellos LF, Cree B, Oksenberg JR, Hauser SL (2007) Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med 357:851-862

34. Aulchenko YS, Hoppenbrouwers IA, Ramagopalan SV, Broer L, Jafari N, Hillert J, Link J, Lundstrom W, Greiner E, Dessa Sadovnick A, Goossens D, Van Broeckhoven C, Del-Favero J, Ebers GC, Oostra BA, van Duijn CM, Hintzen RQ (2008) Genetic variation in the KIF1B locus influences susceptibility to multiple sclerosis. Nat Genet 40:1402-1403

35. Comabella M, Craig DW, Camina-Tato M, Morcillo C, Lopez C, Navarro A, Rio J, Montalban X, Martin R (2008) Identification of a novel risk locus for multiple sclerosis at 13q31.3 by a pooled genome-wide scan of 500,000 single nucleotide polymorphisms. PLoS ONE 3:e3490

36. Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, Radue EW, et al. (2009) Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Hum Mol Genet 18:767-778

37. Byun E, Caillier SJ, Montalban X, Villoslada P, Fernandez O, Brassat D, Comabella M, Wang J, Barcellos LF, Baranzini SE, Oksenberg JR (2008) Genome-wide pharmacogenomic analysis of the response to interferon beta therapy in multiple sclerosis. Arch Neurol 65:337-344

38. Cervino AC, Tsinoremas NF, Hoffman RW (2007) A genome-wide study of lupus: preliminary analysis and data release. Ann N Y Acad Sci 1110:131-139

39. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, Gunnarsson I, Svenungsson E, Sturfelt G, Jonsen A, Truedsson L, Pons-Estel BA, Witte T, D'Alfonso S, Barizzone N, Danieli MG, Gutierrez C, Suarez A, Junker P, Laustrup H, Gonzalez-Escribano MF, Martin J, Abderrahim H, Alarcon-Riquelme ME (2008) Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nat Genet 40:211-216

40. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, Lee AT, et al. (2008) Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med 358:900-909

41. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, Burtt NP, Guiducci C, Parkin M, Gates C, Plenge RM, Behrens TW, Wither JE, Rioux JD, Fortin PR, Graham DC, Wong AK, Vyse TJ, Daly MJ, Altshuler D, Moser KL, Gaffney PM (2008) Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nat Genet 40:1059-1061

42. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, et al. (2008) Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet 40:204-210

43. Jacobson DL, Gange SJ, Rose NR, Graham NM (1997) Epidemiology and estimated population burden of selected autoimmune diseases in the United States. Clin Immunol Immunopathol 84:223-243

44. Van Loveren H, Vos JG, Germolec D, Simeonova PP, Eijkemanns G, McMichael AJ (2001) Epidemiologic associations between occupational and environmental exposures and autoimmune disease: report of a meeting to explore current evidence and identify research needs. Int J Hyg Environ Health 203:483-495

45. Cooper GS, Stroehla BC (2003) The epidemiology of autoimmune diseases. Autoimmun Rev 2:119-125

46. McCarty DJ, Manzi S, Medsger TA, Jr., Ramsey-Goldman R, LaPorte RE, Kwoh CK (1995) Incidence of systemic lupus erythematosus. Race and gender differences. Arthritis Rheum 38:1260-1270

47. Mahid SS, Mulhall AM, Gholson RD, Eichenberger MR, Galandiuk S (2008) Inflammatory bowel disease and African Americans: A systematic review. Inflamm Bowel Dis

48. Rosati G (2001) The prevalence of multiple sclerosis in the world: an update. Neurol Sci 22:117-139

49. Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, Vyse TJ, Rioux JD (2008) Defining the role of the MHC in autoimmunity: a review and pooled analysis. PLoS Genet 4:e1000024

50. Velaga MR, Wilson V, Jennings CE, Owen CJ, Herington S, Donaldson PT, Ball SG, James RA, Quinton R, Perros P, Pearce SH (2004) The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. J Clin Endocrinol Metab 89:5862-5865

51. Smith RL, Warren RB, Eyre S, Ke X, Young HS, Allen M, Strachan D, McArdle W, Gittins MP, Barker JN, Griffiths CE, Worthington J (2008) Polymorphisms in the PTPN22 region are associated with psoriasis of early onset. Br J Dermatol 158:962-968

52. Lee YH, Rho YH, Choi SJ, Ji JD, Song GG, Nath SK, Harley JB (2007) The PTPN22 C1858T functional polymorphism and autoimmune diseases--a meta-analysis. Rheumatology (Oxford) 46:49-56

53. Santin I, Castellanos-Rubio A, Aransay AM, Castano L, Vitoria JC, Bilbao JR (2008) The functional R620W variant of the PTPN22 gene is associated with celiac disease. Tissue Antigens 71:247-249

54. Wang L, Lin S, Rammohan KW, Liu Z, Liu JQ, Liu RH, Guinther N, Lima J, Zhou Q, Wang T, Zheng X, Birmingham DJ, Rovin BH, Hebert LA, Wu Y, Lynn DJ, Cooke G, Yu CY, Zheng P, Liu Y (2007) A dinucleotide deletion in CD24 confers protection against autoimmune diseases. PLoS Genet 3:e49

55. Fung E, Smyth DJ, Howson JM, Cooper JD, Walker NM, Stevens H, Wicker LS, Todd JA (2008) Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. Genes Immun

56. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, Pe'er I, et al. (2007) Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 39:1477-1482

57. Sigurdsson S, Nordmark G, Goring HH, Lindroos K, Wiman AC, Sturfelt G, Jonsen A, Rantapaa-Dahlqvist S, Moller B, Kere J, Koskenmies S, Widen E, Eloranta ML, Julkunen H, Kristjansdottir H, Steinsson K, Alm G, Ronnblom L, Syvanen AC (2005) Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. Am J Hum Genet 76:528-537

58. Miceli-Richard C, Comets E, Loiseau P, Puechal X, Hachulla E, Mariette X (2007) Association of an IRF5 gene functional polymorphism with Sjogren's syndrome. Arthritis Rheum 56:3989-3994

59. Sigurdsson S, Padyukov L, Kurreeman FA, Liljedahl U, Wiman AC, Alfredsson L, Toes R, Ronnelid J, Klareskog L, Huizinga TW, Alm G, Syvanen AC, Ronnblom L (2007) Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis. Arthritis Rheum 56:2202-2210

60. Maas K, Chan S, Parker J, Slater A, Moore J, Olsen N, Aune TM (2002) Cutting edge: molecular portrait of human autoimmune disease. J Immunol 169:5-9

61. Baumgart DC, Carding SR (2007) Inflammatory bowel disease: cause and immunobiology. Lancet 369:1627-1640

62. Baumgart DC, Sandborn WJ (2007) Inflammatory bowel disease: clinical aspects and established and evolving therapies. Lancet 369:1641-1657

63. Orholm M, Binder V, Sorensen TI, Rasmussen LP, Kyvik KO (2000) Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. Scand J Gastroenterol 35:1075-1081

64. Hampe J, Cuthbert A, Croucher PJ, Mirza MM, Mascheretti S, Fisher S, Frenzel H, King K, Hasselmeyer A, MacPherson AJ, Bridger S, van Deventer S, Forbes A, Nikolaus S, Lennard-Jones JE, Foelsch UR, Krawczak M, Lewis C, Schreiber S, Mathew CG (2001) Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. Lancet 357:1925-1928

65. Lesage S, Zouali H, Cezard JP, Colombel JF, Belaiche J, Almer S, Tysk C, O'Morain C, Gassull M, Binder V, Finkel Y, Modigliani R, Gower-Rousseau C, Macry J, Merlin F, Chamaillard M, Jannot AS, Thomas G, Hugot JP (2002) CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. Am J Hum Genet 70:845-857

66. Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, Delmonte T, et al. (2001) Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet 29:223-228

67. Giallourakis C, Stoll M, Miller K, Hampe J, Lander ES, Daly MJ, Schreiber S, Rioux JD (2003) IBD5 is a general risk factor for inflammatory bowel disease: replication of association with Crohn disease and identification of a novel association with ulcerative colitis. Am J Hum Genet 73:205-211

68. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, Mayr G, De La Vega FM, Briggs J, Gunther S, Prescott NJ, Onnie CM, Hasler R, Sipos B, Folsch UR, Lengauer T, Platzer M, Mathew CG, Krawczak M, Schreiber S (2007) A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet 39:207-211

69. Compston A (1997) Genetic epidemiology of multiple sclerosis. J Neurol Neurosurg Psychiatry 62:553-561

70. (1981) Multiple sclerosis in the Orkney and Shetland Islands and in north-east Scotland. Br Med J (Clin Res Ed) 282:502-504

71. Willer CJ, Dyment DA, Risch NJ, Sadovnick AD, Ebers GC (2003) Twin concordance and sibling recurrence rates in multiple sclerosis. Proc Natl Acad Sci U S A 100:12877-12882

72. Kalman B, Lublin FD (1999) The genetics of multiple sclerosis. A review. Biomed Pharmacother 53:358-370

73. Sawcer S, Ban M, Maranian M, Yeo TW, Compston A, Kirby A, Daly MJ, et al. (2005) A high-density screen for linkage in multiple sclerosis. Am J Hum Genet 77:454-467

74. Lincoln MR, Montpetit A, Cader MZ, Saarela J, Dyment DA, Tiislar M, Ferretti V, Tienari PJ, Sadovnick AD, Peltonen L, Ebers GC, Hudson TJ (2005) A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis. Nat Genet 37:1108-1112

75. Saarela J, Kallio SP, Chen D, Montpetit A, Jokiaho A, Choi E, Asselta R, Bronnikov D, Lincoln MR, Sadovnick AD, Tienari PJ, Koivisto K, Palotie A, Ebers GC, Hudson TJ, Peltonen L (2006) PRKCA and multiple sclerosis: association in two independent populations. PLoS Genet 2:e42

76. Barton A, Woolmore JA, Ward D, Eyre S, Hinks A, Ollier WE, Strange RC, Fryer AA, John S, Hawkins CP, Worthington J (2004) Association of protein kinase C alpha (PRKCA) gene with multiple sclerosis in a UK population. Brain 127:1717-1722

77. Gregory SG, Schmidt S, Seth P, Oksenberg JR, Hart J, Prokop A, Caillier SJ, Ban M, Goris A, Barcellos LF, Lincoln R, McCauley JL, Sawcer SJ, Compston DA, Dubois B, Hauser SL, Garcia-Blanco MA, Pericak-Vance MA, Haines JL (2007) Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. Nat Genet 39:1083-1091

78. Lundmark F, Duvefelt K, Iacobaeus E, Kockum I, Wallstrom E, Khademi M, Oturai A, Ryder LP, Saarela J, Harbo HF, Celius EG, Salter H, Olsson T, Hillert J (2007) Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. Nat Genet 39:1108-1113

79. Matesanz F, Caro-Maldonado A, Fedetz M, Fernandez O, Milne RL, Guerrero M, Delgado C, Alcina A (2007) IL2RA/CD25 polymorphisms contribute to multiple sclerosis susceptibility. J Neurol 254:682-684

80. Weber F, Fontaine B, Cournu-Rebeix I, Kroner A, Knop M, Lutz S, Muller-Sarnowski F, Uhr M, Bettecken T, Kohli M, Ripke S, Ising M, Rieckmann P, Brassat D, Semana G, Babron MC, Mrejen S, Gout C, Lyon-Caen O, Yaouanq J, Edan G, Clanet M, Holsboer F, Clerget-Darpoux F, Muller-Myhsok B (2008) IL2RA and IL7RA genes confer susceptibility for multiple sclerosis in two independent European populations. Genes Immun 9:259-263

81. Petri M (2002) Epidemiology of systemic lupus erythematosus. Best Pract Res Clin Rheumatol 16:847-858

82. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, Schaller JG, Talal N, Winchester RJ (1982) The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 25:1271-1277

83. Tsao BP (2003) The genetics of human systemic lupus erythematosus. Trends Immunol 24:595-602

84. Edberg JC, Langefeld CD, Wu J, Moser KL, Kaufman KM, Kelly J, Bansal V, Brown WM, Salmon JE, Rich SS, Harley JB, Kimberly RP (2002) Genetic linkage and association of Fcgamma receptor IIIA (CD16A) on chromosome 1q23 with human systemic lupus erythematosus. Arthritis Rheum 46:2132-2140

85. Prokunina L, Castillejo-Lopez C, Oberg F, Gunnarsson I, Berg L, Magnusson V, Brookes AJ, Tentler D, Kristjansdottir H, Grondal G, Bolstad AI, Svenungsson E, Lundberg I, Sturfelt G, Jonssen A, Truedsson L, Lima G, Alcocer-Varela J, Jonsson R, Gyllensten UB, Harley JB, Alarcon-Segovia D, Steinsson K, Alarcon-Riquelme ME (2002) A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. Nat Genet 32:666-669

86. Bowman SJ, Ibrahim GH, Holmes G, Hamburger J, Ainsworth JR (2004) Estimating the prevalence among Caucasian women of primary Sjogren's syndrome in two general practices in Birmingham, UK. Scand J Rheumatol 33:39-43

87. Thomas E, Hay EM, Hajeer A, Silman AJ (1998) Sjogren's syndrome: a community-based study of prevalence and impact. Br J Rheumatol 37:1069-1076

88. Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, Daniels TE, Fox PC, Fox RI, Kassan SS, Pillemer SR, Talal N, Weisman MH (2002) Classification criteria for Sjogren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. Ann Rheum Dis 61:554-558

89. Bolstad AI, Jonsson R (2002) Genetic aspects of Sjogren's syndrome. Arthritis Res 4:353-359

90. Hulkkonen J, Pertovaara M, Antonen J, Pasternack A, Hurme M (2001) Elevated interleukin-6 plasma levels are regulated by the promoter region polymorphism of the IL6 gene in primary Sjogren's syndrome and correlate with the clinical manifestations of the disease. Rheumatology (Oxford) 40:656-661

91. Hulkkonen J, Pertovaara M, Antonen J, Lahdenpohja N, Pasternack A, Hurme M (2001) Genetic association between interleukin-10 promoter region polymorphisms and primary Sjogren's syndrome. Arthritis Rheum 44:176-179

92. Korman BD, Alba MI, Le JM, Alevizos I, Smith JA, Nikolov NP, Kastner DL, Remmers EF, Illei GG (2008) Variant form of STAT4 is associated with primary Sjogren's syndrome. Genes Immun 9:267-270

93. Isaacs A, Lindenmann J (1957) Virus interference. I. The interferon. Proc R Soc Lond B Biol Sci 147:258-267

94. Owerbach D, Rutter WJ, Shows TB, Gray P, Goeddel DV, Lawn RM (1981) Leukocyte and fibroblast interferon genes are located on human chromosome 9. Proc Natl Acad Sci U S A 78:3123-3127

95. Baccala R, Kono DH, Theofilopoulos AN (2005) Interferons as pathogenic effectors in autoimmunity. Immunol Rev 204:9-26

96. Ronnblom LE, Alm GV, Oberg KE (1991) Autoimmunity after alpha-interferon therapy for malignant carcinoid tumors. Ann Intern Med 115:178-183

97. Baechler EC, Batliwalla FM, Karypis G, Gaffney PM, Ortmann WA, Espe KJ, Shark KB, Grande WJ, Hughes KM, Kapur V, Gregersen PK, Behrens TW (2003) Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. Proc Natl Acad Sci U S A 100:2610-2615

98. van der Pouw Kraan TC, Wijbrandts CA, van Baarsen LG, Voskuyl AE, Rustenburg F, Baggen JM, Ibrahim SM, Fero M, Dijkmans BA, Tak PP, Verweij CL (2007) Rheumatoid arthritis subtypes identified by genomic profiling of peripheral blood cells: assignment of a type I interferon signature in a subpopulation of patients. Ann Rheum Dis 66:1008-1014

99. Hjelmervik TO, Petersen K, Jonassen I, Jonsson R, Bolstad AI (2005) Gene expression profiling of minor salivary glands clearly distinguishes primary Sjogren's syndrome patients from healthy control subjects. Arthritis Rheum 52:1534-1544

100. van Baarsen LG, van der Pouw Kraan TC, Kragt JJ, Baggen JM, Rustenburg F, Hooper T, Meilof JF, Fero MJ, Dijkstra CD, Polman CH, Verweij CL (2006) A subtype of multiple sclerosis defined by an activated immune defense program. Genes Immun 7:522-531

101. Garcia-Sastre A, Biron CA (2006) Type 1 interferons and the virus-host relationship: a lesson in detente. Science 312:879-882

102. Pestka S, Krause CD, Walter MR (2004) Interferons, interferon-like cytokines, and their receptors. Immunol Rev 202:8-32

103. Fattovich G, Giustina G, Favarato S, Ruol A (1996) A survey of adverse events in 11,241 patients with chronic viral hepatitis treated with alfa interferon. J Hepatol 24:38-47

104. Ronnblom LE, Alm GV, Oberg KE (1990) Possible induction of systemic lupus erythematosus by interferon-alpha treatment in a patient with a malignant carcinoid tumour. J Intern Med 227:207-210

105. Fitzgerald-Bocarsly P (2002) Natural interferon-alpha producing cells: the plasmacytoid dendritic cells. Biotechniques Suppl:16-20, 22, 24-19

106. Blomberg S, Eloranta ML, Cederblad B, Nordlin K, Alm GV, Ronnblom L (2001) Presence of cutaneous interferon-alpha producing cells in patients with systemic lupus erythematosus. Lupus 10:484-490

107. Rock FL, Hardiman G, Timans JC, Kastelein RA, Bazan JF (1998) A family of human receptors structurally related to Drosophila Toll. Proc Natl Acad Sci U S A 95:588-593

108. Kyogoku C, Tsuchiya N (2007) A compass that points to lupus: genetic studies on type I interferon pathway. Genes Immun 8:445-455

109. Bauer S, Pigisch S, Hangel D, Kaufmann A, Hamm S (2008) Recognition of nucleic acid and nucleic acid analogs by Toll-like receptors 7, 8 and 9. Immunobiology 213:315-328

110. Lovgren T, Eloranta ML, Bave U, Alm GV, Ronnblom L (2004) Induction of interferon-alpha production in plasmacytoid dendritic cells by immune complexes containing nucleic acid released by necrotic or late apoptotic cells and lupus IgG. Arthritis Rheum 50:1861-1872

111. Lovgren T, Eloranta ML, Kastner B, Wahren-Herlenius M, Alm GV, Ronnblom L (2006) Induction of interferon-alpha by immune complexes or liposomes containing systemic lupus erythematosus autoantigen- and Sjogren's syndrome autoantigen-associated RNA. Arthritis Rheum 54:1917-1927

112. Ronnblom L, Eloranta ML, Alm GV (2006) The type I interferon system in systemic lupus erythematosus. Arthritis Rheum 54:408-420

113. Bave U, Nordmark G, Lovgren T, Ronnelid J, Cajander S, Eloranta ML, Alm GV, Ronnblom L (2005) Activation of the type I interferon system in primary Sjogren's syndrome: a possible etiopathogenic mechanism. Arthritis Rheum 52:1185-1195

114. Taniguchi T, Ogasawara K, Takaoka A, Tanaka N (2001) IRF family of transcription factors as regulators of host defense. Annu Rev Immunol 19:623-655

115. Barnes BJ, Moore PA, Pitha PM (2001) Virus-specific activation of a novel interferon regulatory factor, IRF-5, results in the induction of distinct interferon alpha genes. J Biol Chem 276:23382-23390

116. Takaoka A, Yanai H, Kondo S, Duncan G, Negishi H, Mizutani T, Kano S, Honda K, Ohba Y, Mak TW, Taniguchi T (2005) Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. Nature 434:243-249

117. Yanai H, Chen HM, Inuzuka T, Kondo S, Mak TW, Takaoka A, Honda K, Taniguchi T (2007) Role of IFN regulatory factor 5 transcription factor in anti-viral immunity and tumor suppression. Proc Natl Acad Sci U S A 104:3402-3407

118. Hu G, Barnes BJ (2009) IRF-5 Is a Mediator of the Death Receptor-induced Apoptotic Signaling Pathway. J Biol Chem 284:2767-2777

119. Mancl ME, Hu G, Sangster-Guity N, Olshalsky SL, Hoops K, Fitzgerald-Bocarsly P, Pitha PM, Pinder K, Barnes BJ (2005) Two discrete promoters regulate the alternatively spliced human interferon regulatory factor-5 isoforms. Multiple isoforms with distinct cell type-specific expression, localization, regulation, and function. J Biol Chem 280:21078-21090

120. Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, Johnson KP, Sibley WA, Silberberg DH, Tourtellotte WW (1983) New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol 13:227-231

121. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, Lublin FD, Metz LM, McFarland HF, O'Connor PW, Sandberg-Wollheim M, Thompson AJ, Weinshenker BG, Wolinsky JS (2005) Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Ann Neurol 58:840-846

122. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, Liew A, et al. (2007) TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. N Engl J Med 357:1199-1209

123. Syvanen AC, Aalto-Setala K, Harju L, Kontula K, Soderlund H (1990) A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. Genomics 8:684-692

124. Gustafsdottir SM, Schlingemann J, Rada-Iglesias A, Schallmeiner E, Kamali-Moghaddam M, Wadelius C, Landegren U (2007) In vitro analysis of DNA-protein interactions by proximity ligation. Proc Natl Acad Sci U S A 104:3067-3072

125. Blangero J, Goring HH, Kent JW, Jr., Williams JT, Peterson CP, Almasy L, Dyer TD (2005) Quantitative trait nucleotide analysis using Bayesian model selection. Hum Biol 77:541-559

126. Nyholt DR (2006) ssSNPer: identifying statistically similar SNPs to aid interpretation of genetic association studies. Bioinformatics 22:2960-2961

127. Sigurdsson S, Nordmark G, Garnier S, Grundberg E, Kwan T, Nilsson O, Eloranta ML, Gunnarsson I, Svenungsson E, Sturfelt G, Bengtsson AA, Jonsen A, Truedsson L, Rantapaa-Dahlqvist S, Eriksson C, Alm G, Goring HH, Pastinen T, Syvanen AC, Ronnblom L (2008) A risk haplotype of STAT4 for systemic lupus erythematosus is over-expressed, correlates with anti-dsDNA and shows additive effects with two risk alleles of IRF5. Hum Mol Genet 17:2868-2876

128. Martinez A, Varade J, Marquez A, Cenit MC, Espino L, Perdigones N, Santiago JL, Fernandez-Arquero M, de la Calle H, Arroyo R, Mendoza JL, Fernandez-Gutierrez B, de la Concha EG, Urcelay E (2008) Association of the STAT4 gene with increased susceptibility for some immune-mediated diseases. Arthritis Rheum 58:2598-2602

129. Das S, Ward SV, Tacke RS, Suske G, Samuel CE (2006) Activation of the RNA-dependent protein kinase PKR promoter in the absence of interferon is dependent upon Sp proteins. J Biol Chem 281:3244-3253

130. Graham RR, Kozyrev SV, Baechler EC, Reddy MV, Plenge RM, Bauer JW, Ortmann WA, Koeuth T, Gonzalez Escribano MF, Pons-Estel B, Petri M, Daly M, Gregersen PK, Martin J, Altshuler D, Behrens TW, Alarcon-Riquelme ME (2006) A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. Nat Genet 38:550-555

131. Graham RR, Kyogoku C, Sigurdsson S, Vlasova IA, Davies LR, Baechler EC, Plenge RM, et al. (2007) Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. Proc Natl Acad Sci U S A 104:6758-6763

132. Kozyrev SV, Lewen S, Reddy PM, Pons-Estel B, Witte T, Junker P, Laustrup H, Gutierrez C, Suarez A, Francisca Gonzalez-Escribano M, Martin J, Alarcon-Riquelme ME (2007) Structural insertion/deletion variation in IRF5 is associated with a risk haplotype and defines the precise IRF5 isoforms expressed in systemic lupus erythematosus. Arthritis Rheum 56:1234-1241

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 434

Editor: The Dean of the Faculty of Medicine