Thesis for doctoral degree (Ph.D.) 2009

Thesis for doctoral degree (Ph.D.) 2009

SUSCEPTIBILITY GENES IN SYSTEMIC LUPUS ERYTHEMATOSUS

Anna Hellquist

Thesis for doctoral degree (Ph.D.)
2009

SUSCEPTIBILITY GENES IN SYSTEMIC LUPUS ERYTHEMATOSUS

Anna Hellquist

KAROLINSKA INSTITUTET ANNO 1810
Karolinska Institutet

KAROLINSKA INSTITUTET ANNO 1810
Karolinska Institutet

From DEPARTMENT OF BIOSCIENCE AND NUTRITION
Karolinska Institutet, Stockholm, Sweden

SUSCEPTIBILITY GENES IN SYSTEMIC LUPUS ERYTHEMATOSUS

Anna Hellquist

Karolinska Institutet

Stockholm 2009

Published by Karolinska Institutet.

ISBN 978-91-7409-660-6

Printed by
REPROPRINT AB
Stockholm 2009
www.reproprint.se
Gårdsvägen 4, 169 70 Solna

Lupus, is it lupus?
(George Costanza)

It's never lupus!
(Dr. Gregory House)

ABSTRACT

Systemic lupus erythematosus (SLE) is a complex autoimmune disease for which the incidence and prevalence vary between populations and also between males and females. SLE is characterized by production of pathogenic autoantibodies against nuclear antigens due to a breakdown in self-tolerance and the pathogenesis is associated with the formation of immune complexes, followed by tissue inflammation in multiple organs, such as the skin, joints, heart and kidneys. SLE is an unusually heterogeneous disease and its clinical classification is based on criteria set by the American College of Rheumatology (ACR). Although the underlying pathogenic mechanisms of SLE remain imperfectly understood, both environmental influences and genetic factors have been found to play an important role in disease initiation and progression. Both familial aggregation studies and twin studies support a strong genetic component in SLE and today over 30 convincingly associated SLE susceptibility genes have been identified. Many of these SLE-predisposing genes appear to be involved in similar and/or related biological pathways, including the processing of immune complexes, type I interferon production, and immune signal transduction. Other genes, on the contrary, have no assigned function or obvious role in the immune system, and thus represent ideal candidate to reveal novel disease mechanisms.

The aim of this thesis was to study susceptibility genes in SLE, using a number of different approaches. In Paper I we performed a functional candidate gene association study of the *GIMAP5* gene, which had been shown to be essential for the survival of leukocytes, and identified association between this gene and SLE in two independent family cohorts from Finland and the UK. In Paper II we performed a positional mapping study within our previously identified susceptibility loci on chromosomes 14q21-q23 and identified association to the novel SLE candidate gene *MAMDC1* in four independent cohorts. This gene appears to encode for a novel member of the immunoglobulin cell adhesion molecules superfamily, which is involved in cell adhesion, migration, and recruitment to inflammatory sites. In Papers III-V we performed three different replication studies of previously identified SLE susceptibility in a Finnish case-control cohort and identified association to several genes, including *STAT4*, *IRF5-TNPO3*, *TYK2*, *ITGAM-ITGAX*, *TNFAIP3*, *FAM167A-BLK*, *BANK1* and *KIAA1542*. We furthermore showed evidence of gene-gene interaction (epistasis) between SNPs in *IRF5* and *TYK2*.

In conclusion we have identified two novel SLE candidate genes contributing to SLE susceptibility in several populations as well as shown that a number of previously identified SLE susceptibility genes also contribute to risk in the Finnish population.

LIST OF PUBLICATIONS

I. **Hellquist A***, Zucchelli M*, Kivinen K, Saarialho-Kere U, Koskenmies S, Widén E, Julkunen H, Wong A, Karjalainen-Lindsberg MJ, Skoog T, Vendelin J, Cunninghame Graham DS, Vyse TJ, Kere J and Lindgren CM.
The human *GIMAP5* gene has a common polyadenylation polymorphism increasing risk to systemic lupus erythematosus.
Journal of Medical Genetics 2007; May 44(5): 314-21.

II. **Hellquist A**, Zucchelli M, Lindgren CM, Saarialho-Kere U, Järvinen TM, Koskenmies S, Julkunen H, Onkamo P, Skoog T, Panelius J, Räisänen-Sokolowski A, Hasan T, Widén E, Gunnarson I, Svenungsson E, Padyukov L, Assadi G, Berglind L, Mäkelä V, Kivinen K, Wong A, Cunninghame Graham DS, Vyse TJ, D'Amato M and Kere J.
Identification of *MAMDC1* as a candidate susceptibility gene for systemic lupus erythematosus (SLE).
Manuscript accepted in PLoS ONE

III. **Hellquist A***, Järvinen TM*, Koskenmies S, Zucchelli M, Orsmark-Pietras C, Berglind L, Panelius J, Hasan T, Julkunen H, D'Amato M, Saarialho-Kere U and Kere J.
Evidence for genetic association and interaction between the *TYK2* and *IRF5* genes in systemic lupus erythematosus.
The Journal of Rheumatology 2009; 36(8): 1631-8.

IV. **Hellquist A***, Sandling J*, Zucchelli M, Koskenmies S, Julkunen H, D'Amato M, Garnier S, Syvänen AC and Kere J.
Variation in *STAT4* is associated with systemic lupus erythematosus (SLE) in a Finnish family cohort.
Annals of the Rheumatic Diseases 2009; Aug 27 (PMID: 19717398).

V. **Hellquist A**, Järvinen TM, Zucchelli M, Koskenmies S, Julkunen H, D'Amato M, Saarialho-Kere U and Kere J.
Investigation of recently identified SLE genome-wideassociation genes reveals the strongest association to *STAT4*, *IRF5* and *ITGAM* in the Finnish population.
In manuscript

* Authors contributed equally

CONTENTS

LIST OF ABBREVIATIONS

ACR	American College of Rheumatology
ANA	Anti-nuclear autoantibody
APC	Antigen-presenting cell
BCR	B-cell receptor
bp	Base pairs
CAM	Cell adhesion molecule
cDNA	Complementary deoxyribonucleic acid
CI	Confidence interval
cM	Centimorgan
CNS	Central nervous system
CNV	Copy number variant
C_t	Threshold cycle
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EBV	Epstein-Barr virus
FNIII	Fibronectin type III
HLA	Human leukocyte antigen
HPM	Haplotype pattern mining
HWE	Hardy-Weinberg equilibrium
kb	Kilobase
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
Indel	Insertion – deletion
LD	Linkage disequilibrium
LOD	Logarithm of odds
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
MAM	Meprin/A5-protein/PTPmu
Mb	Megabase
MHC	Major histocompatibility complex
mRNA	Messenger ribonucleic acid
NSAID	Non-steroidal anti-inflammatory drug
OR	Odds ratio
PCR	Polymerase chain reaction
PDT	Pedigree disequilibrium test
PolyA	Polyadenylation
R	Receptor
RR	Risk ratio
RA	Rheumatoid arthritis
RNA	Ribonucleic acid

RT-PCR	Real-time polymerase chain reaction
pDC	Plasmacytoid dendritic cells
qRT-PCR	Quantitative real-time polymerase chain reaction
SES	Socioeconomic status
SNP	Single nucleotide polymorphism
SLE	Systemic lupus erythematosus
SSR	Simple sequence repeats
T1D	Type 1 diabetes
TGF	Transforming growth factor
TLR	Toll-like receptor
TNF	Tumour necrosis factor
TDT	Transmission disequilibrium test
Th	T helper cell
T reg	T regulatory cell
UTR	Untranslated region
UV	Ultraviolet light

1 POPULÄRVETENSKAPLIG SAMMANFATTNING

Jag har många gånger fått frågor från både släkt och vänner om vad det egentligen är jag sysslar med på Karolinska Institutet och jag inser att mina förklaringar kanske inte alltid har nått hela vägen fram. Därför vill jag under den här rubriken förklara den forskning jag ägnat mig åt under de senaste sex åren på ett sätt som kan förstås av vem som helst, inte bara mina närmsta forskningskollegor. Säkert känner du någon med en sjukdom där immunförsvaret, utöver sin normala funktion som är att skydda oss från bakterie- och virusinfektioner, har löpt amok. Det vanligaste är att immunförsvaret börjar bekämpa helt ofarliga ämnen, som t.ex. pollen, och därmed ger upphov till allergi. Men det händer även att kroppens egna friska celler attackeras, vilket leder till inflammation. Exempelvis så angrips en viss typ av celler i bukspottskörteln vid typ 1 diabetes (barndiabetes) eller nervcellerna i centrala nervsystemet vid multipel skleros. Den här typen av sjukdomar kallas för autoimmuna sjukdomar.

Vid SLE (systemisk lupus erythematosus), som är den sjukdom jag studerat, så angrips vissa komponenter som finns i alla kroppens celler. Detta leder till att en mängd olika vävnader och organ kan drabbas, vilket i sin tur gör att två individer med SLE kan ha väldigt olika symptom. Det finns dock en del allmänna symptom, bland annat onormal trötthet, som en stor del av patienterna är drabbade av. Även hudutslag också vanligt och då främst på de delar av kroppen som utsätts för sol. Vissa patienter får utslag i ansiktet i form av en fjäril (butterfly rash), vilket har gjort att SLE ofta symboliseras med just en fjäril och som också är tanken bakom omslaget till denna avhandling. Andra organ som ofta drabbas av inflammation vid SLE är lungorna, hjärtat och njurarna. Det förekommer även psykiska symptom som depressioner och inlärningssvårigheter. Denna bredd av symptom har lett till att man har utvecklat ett system för att avgöra om en patient har SLE och enligt det systemet måste en patient visa fyra utav de elva olika sjukdomssymptom som är typiska för SLE.

Vad som är vanligt för de flesta autoimmuna sjukdomarna, och så också för SLE, är att de går i skov. Skov innebär en tillfällig försämring av sjukdomsförloppet, vilket också betyder att man kan känna sig nästan helt frisk i perioder. Vid lindriga SLE skov behandlas patienterna med låga doser av kortison som är ett läkemedel som dämpar immunförsvaret och därmed den akuta inflammationen som uppstår vid ett skov. Vid svåra skov använder man sig av höga doser av kortison, parallellt med att man också behandlar med cytostatika, vilken motverkar att immuncellerna delar sig och därmed blir deras förmåga att attackera kroppen mindre.

Varför en person drabbas av SLE vet man inte riktigt och det är här vi forskare kommer in i bilden. Det vi vill med våra undersökningar är att förståelsen om de sjukdomsmekanismer som ligger bakom SLE ökar. Detta kan i sin tur leda till att effektivare behandlingsmetoder, och på sikt kanske till och med att förebyggande behandling kan utvecklas. Vad vet vi då i nuläget? Jo för det första så vet vi att våra gener fyller en viktig roll när det gäller en individs risk att utveckla SLE (eller någon

annan autoimmun sjukdom för den delen). Detta vet man bland annat genom att man har studerat förekomsten av SLE hos tvillingar och sett att hos enäggstvillingar (som är genetiskt identiska) är det tio gånger vanligare att båda tvillingarna har SLE, jämfört med tvääggstvillingar. Men för att drabbas av SLE räcker det inte bara med att ha ett genetiskt anlag utan även vissa miljöfaktorer, där virusinfektioner, UV-strålning och rökning är några av de faktorer som misstänks spela en roll, är nödvändiga för att utlösa sjukdomen. Vi vet också att en enskild gen och miljöfaktor påverkar risken att insjukna med väldigt lite och det är med stor sannolikhet en mängd gener i samverkan med ett antal olika miljöfaktorer som bidrar till att en individ blir sjuk. Vad som är den avgörande faktorn till att man faktiskt blir sjuk är dock ännu inte klarlagt. Detta behöver inte heller vara samma faktor för alla individer.

Detta komplexa scenario med både arv- och miljöpåverkan, där varje enskild faktor har relativt liten effekt, har gjort att identifieringen av de gener som orsakar SLE har tagit lång tid. År 2006 kom en teknik som har gjort det möjligt att titta på ett stort antal genetiska variationer på samma gång, något som tidigare har krävt extremt mycket tid och pengar. Denna typ av analys kallas ”genome-wide association” studie, vilket brukar förkortas GWA studie, och den innebär att hela arvsmassan skannas av i ett stort antal sjuka och friska för att hitta sjukdomskopplade variationer. Lite förenklat så är alla typer av genetiska studier för att identifiera sjukdomsgener baserade på en jämförelse av skillnader i arvsanlaget mellan sjuka och friska. Som de flesta säkert vet så utgörs vår arvsmassa av DNA som är en dubbelsträngad kedja bestående av fyra baser - A, C, G och T. Till största delen är denna kedja identisk hos alla människor, men vid ungefär var tusende bas så har man kunnat visa att den skiljer sig åt. En individ kan till exempel ha basen A på precis samma ställe där en annan individ har basen G. I vissa fall så kan denna lilla förändring faktiskt vara skillnaden mellan sjuk och frisk, vilket är just vad de genetiska studierna tittar på. Om vi exempelvis identifierar basen A hos ett antal sjuka individer, medan vi hos friska individer, på exakt samma ställe i texten identifierar ett G, så betyder det att det troligtvis finns en koppling mellan basen A och sjukdomen. Vad den sjukdomskopplade basen orsakar för förändring i kroppen är väldigt olika, men generellt kan man säga att den har ganska lite effekt i sig självt då även en stor del friska individer bär omkring på samma variant. Det är dock inte alltid så att den sjukdomskopplade varianten befinner sig inuti en gen, vilka är de platser i arvsmassa som ger upphov, eller kodar för, de proteiner som bygger upp vår kropp. Detta betyder dock inte att de är oviktiga men vilken roll de har är svårare att studera.

I en GWA studie så kan man undersöka flera hundratusentals sådana variationer hos tusentals individer och det har lett till att totalt ungefär 30 gener som bidrar till att en individ får SLE har identifierats de senaste åren (dessa finns listade i tabell 2 i avhandlingen). Då jag började mina doktorandstudier så fanns inte denna teknik tillgänglig och antalet identifierade gener som man säkert kunde säga att de orsakade SLE var endast ett fåtal. De vanligaste metoderna man använde sig av för att hitta sjukdomsgener var antingen att 1) jämföra nedärvningsmönster av specifika markörer i familjer med SLE (kopplingsanalys) eller 2) jämföra ett fåtal variationer mellan sjuka och friska (associationsanalys). En associationsanalys bygger på samma princip som en

GWA studie, fast man tittar på ett begränsat område i arvsmassan eller i en specifik gen som man misstänkte kunde vara inblandad i SLE. Den stora skillnaden är att i en GWA studie begränsas man inte av en sådan hypotes.

I min avhandling har jag använt mig av associationsanalys och på så sätt kunnat identifiera två gener som bidrar till att en individ utvecklar SLE. Det DNA vi främst har använt oss av för att hitta dessa gener kommer från sjuka och friska familjemedlemmar i totalt 192 familjer från Finland. För att vi ska kunna vara säkra på våra upptäckter så har vi även tittat på dessa gener i flera andra oberoende material med SLE patienter och friska individer med ursprung i England, Sverige och Finland. Den ena av de gener vi hittat, *MAMDC1*, är tämligen ostuderad och vi vet ännu inte vad den har för roll i SLE. Vad vi vet är att den ger upphov till ett protein vars funktion är att hjälpa celler att binda till varandra, vilket är viktigt för en rad immunfunktioner. Den andra genen, *GIMAP5*, är något mer studerad och vi vet att den är viktig för överlevnaden av de vita blodkropparna. Detta vet vi för att möss och råttor med en mutation i den här genen har en onormalt låg nivå av vita blodkroppar, något som också förekommer hos SLE patienter. Men inte heller här vet vi exakt vad som går fel vid SLE, då vi hos våra patienter inte hittat samma mutation som hos möss och råttor. Men vi har sett att våra patienter verkar uttrycka en lite annorlunda form av *GIMAP5* genen jämfört med friska individer, vilket vi tror är förklaringen till att individer med den variationen lättare får SLE. I de tre andra studierna som ingår i min avhandling har vi valt en lite annan väg, vilken har varit att undersöka om ett antal av de gener som andra forskargrupper har hittat också bidrar till SLE i våra finska patienter. Som resultat av dessa tre studier så vet vi nu också att generna *STAT4*, *IRF5*, *ITGAM*, *TNFAIP3*, *FAM167A-BLK*, *BANK1*, *KIAA1542* och *TYK2* också är riskfaktorer hos våra patienter utöver de två gener som vår grupp själva har identifierat.

Vad som är viktigt med att studera en sjukdom som SLE är att vi får ledtrådar till vad som har gått fel även vid andra autoimmuna sjukdomar. Det har nämligen visat sig att många av de riskgener som har identifierats i SLE under de senaste åren är gemensamma mellan en rad sjukdomar (och tvärt om). Detta har lett till att man misstänker att det finns en rad gener som är viktiga när det gäller att behålla toleransen mot den egna vävnaden, vilket är centralt för alla autoimmuna sjukdomar. Utöver dessa finns sedan en rad gener som verkar vara specifika för respektive sjukdom och som bestämmer vilken sjukdom som just den personen får.

Trots att det fortfarande är en lång väg kvar till dess att vi till fullo förstår vad som går fel i vårt immunförsvar vid SLE har vi under de senaste åren kommit en bra bit på väg. Det har även framkommit vilken oerhörd styrka det finns i att utföra genetiska studier då det vi studerar är de faktiska felen och inte följden av själva sjukdomsprocessen. Eller som jag såg det beskrivet i en artikel – genetiska studier är som att undersöka ett maskinhaveri genom att gå igenom ritningarna och leta efter fel i designen, istället för att för att försöka bygga ihop alla tusentals vrakdelar från den kollapsade maskinen och på så sätt lista ut vad som gick fel. Det är en väldigt träffande beskrivning om vad vi genetiker egentligen sysslar med.

2 BACKGROUND

2.1 SYSTEMIC LUPUS ERYTHEMATOSUS

2.1.1 The history of SLE

The disease systemic lupus erythematosus (SLE) has been known to exist for over a thousand years and the word lupus (wolf in Latin) was first used in this context as early as the 10th century, most likely because the destructive cutaneous (skin) injuries caused by the disease reminded of the bites of a wolf (reviewed in (Mallavarapu and Grimsley, 2007)). The term systemic lupus erythematosus was coined in 1895, while the modern period of understanding this disease began around 1950, with observations leading to the identification of SLE as an autoimmune disease.

2.1.2 General aspects of SLE susceptibility

SLE is commonly called the prototype of complex autoimmune diseases and is characterized by production of pathogenic autoantibodies against various nuclear antigens due to a breakdown in self-tolerance. This results in a wide range of immunologic abnormalities and immune complex formation which subsequently leads to multiple tissue and organ damages. As a result, SLE is a heterogeneous disorder that affects individuals with a wide range of clinical manifestations. Why an individual develops SLE is not completely understood but most likely multiple susceptibility genes interacting with a variety of potential environmental exposures are involved.

2.1.3 Clinical aspects, diagnosis and outcome of SLE

SLE primarily occurs in women and approximately 9 of 10 cases of SLE are females (Masi and Kaslow, 1978) in their childbearing age, with a usual disease onset between ages 15 and 40. The typical SLE patient is a young woman presenting with intermittent fatigue, joint pain and swelling, skin rashes (butterfly rash), low white blood cell count and anemia. Approximately one-half of patients will present with more severe complication, such as nephritis, central nervous system (CNS) vasculitis, pulmonary hypertension, interstitial lung disease, and stroke (Arbuckle et al., 2003). Although a variety of organ systems can be affected in human SLE, targeting of the kidneys is the most severe clinical pathology and many of the clinical manifestations correlating with morbidity and death are associated with renal failure (Balow, 2005). The most characteristic clinical feature of SLE is the production of anti-nuclear autoantibodies (ANAs), which are present in 95% of all cases (Isenberg et al., 2007; Manzi, 2009). However, ANAs also commonly occur in the healthy population and are detected sporadically in up to 2% of the female population over the age of 40 (Wakeland et al., 2001). Other antibodies, such as those directed against anti-double-stranded DNA (dsDNA), are highly specific for SLE, however, not that sensitive given that they are only present in 70% patients with SLE (Isenberg and Collins, 1985). Similar to many autoimmune diseases SLE goes from being non-active to being active, i.e. the disease

course flares. Ultraviolet (UV) light, infection, stress, pregnancies or certain drugs can be possible triggers.

Given that SLE is a heterogeneous disease, its diagnosis is highly complicated and therefore based on the fulfillment of at least four out of eleven classification criteria set by the American College of Rheumatology (ACR) (Hochberg, 1997; Tan et al., 1982), described in Table 1. As a consequence, two individuals diagnosed with SLE can have completely different symptoms and a remarkable 330 diagnostic combinations are theoretically possible (Eisenberg, 2009). However, some of the criteria, including for example positive ANAs, are more common than others, while for example the neurological disorders are less common (Petri, 2005). As a result of this heterogeneity, and the fact that symptoms evolve over time, it takes an average of four years before patients are correctly diagnosed with SLE (Manzi, 2009). On the other hand SLE is also often over-diagnosed. When evaluating individuals previously given a presumptive diagnosis of SLE by a non-rheumatologist only about half of those individuals could be confirmed as having SLE (Narain et al., 2004). Out of the misdiagnosed individuals, about 5% had a different autoimmune disease such as systemic sclerosis or Sjögren's syndrome, 5% had fibromyalgia, 29% had positive ANA but no autoimmune disease and 10% had a non-rheumatic disease.

Table 1. The 1982 Revised Criteria for Classification of Systemic Lupus Erythematosus from the American College of Rheumatology.*

Category		Symptom
Skin criteria	1.	Butterfly (Malar) rash (rash over the cheeks and nose)
	2.	Discoid rash (red, flaking and possibly scarring rashes, predominantly on the face)
	3.	Photosensitivity (rash after exposure to sunlight)
	4.	Oral ulcerations
Systemic criteria	5.	Arthritis (usually pain in the joints)
	6.	Serositis (pleuritis or pericarditis)
	7.	Renal disorders
	8.	Neurological disorders (seizures or psychosis)
Laboratory criteria	9.	Hematologic: haemolytic anemia with reticulocytosis leukopenia, lymphopenia or thrombocytopenia
	10.	Immunologic: positive LE cell preparation, anti-dsDNA antibodies, anti-Sm antibodies and anti-phospholipid antibodies
	11.	Antinuclear antibodies (ANAs)

* Adapted from Immunity, 15(3), Wakeland et al, Delineating the Genetic Basis of Systemic Lupus Erythematosus , 397-408, Copyright (2001), with permission from Elsevier.

In the 1950s, the five-year survival for a newly diagnosed SLE patient was approximately 50% (Merrell and Shulman, 1955). However, with better treatment and earlier diagnosis the survival has improved and the five-year survival rate is today

expected to be 95% for most patients (Lau et al., 2006; Manzi, 2009). Before exogenous corticosteroids and immunosuppressant drugs were introduced as treatment for SLE most patients died of active disease or infection (Manzi, 2009). Although therapies for SLE allow management of disease severity, a variety of deleterious drug side effects and therapy resistant disease symptoms significantly diminish the quality of life for many patients. Heart disease, cancer and osteoporosis are also prominent problems, where in the case of heart disease, both atherosclerotic and subclinical cardiovascular diseases are increased (Manzi, 2009). This does not appear to be a consequence of the traditional risk factors for cardiovascular disease, such as metabolic syndrome or hypertension which are common in SLE, since adjustment for these factors still gives about a 7 to 10 time higher risk for non-fatal coronary heart disease and a 17 time higher risk of fatal coronary heart disease (Esdaile et al., 2001). Also the risk for hematologic and possibly lung and hepatobiliary cancers is increased in SLE (Manzi, 2009). However, smoking may be a confounding factor in terms of cancer risk since SLE is more common in current smokers (see section 2.1.6.2). The increased risk of osteoporosis is partly due to treatment with corticosteroids; however, the risk is still increased after this factor is controlled for (Manzi, 2009).

2.1.4 Therapy of SLE

Treatment of SLE depends on the severity of disease, which ranges from mild to severe. Traditional management of SLE has normally included treatment with non-steroidal anti-inflammatory drugs (NSAIDs) and the anti-malarial drug hydroxychloroquine for mild symptoms or flares, where common manifestations are arthritis, rashes, photosensitivity and fatigue. For intermediate disease, which include manifestations such as serositis, severe rashes and hematological manifestations, as well as for severe disease, categorized by renal, CNS, severe skin or hematological manifestation, the additional use of corticosteroids and non-specific immunosuppressive drugs is required. Several new treatments are now being tested in SLE, including B-cell depleting therapies, antibodies and fusion proteins that block interleukins or the cross-talk between B- and T-cells (reviewed in (Sousa and Isenberg, 2009)).

2.1.5 Incidence and prevalence of SLE

Several studies have looked at the prevalence and incidence of SLE, however, the data are sometimes conflicting and also differ between countries, partially due to differences in study methodology (Danchenko et al., 2006). Danchenko et al. (2006) summarized the results of over 60 studies looking at incidence and prevalence in the USA, Canada, Australia, Japan, Martinique and several European countries. Worldwide, the lowest overall incidence was found in Iceland (3/100,000) and Japan (3/100,000), and the highest in the USA (5/100,000) and France (5/100,000). The overall prevalence was the lowest in Northern Ireland (25/100,000), the UK (26/100,000) and Finland (28/100,000), and the highest in Italy (71/100,000), Spain (91/100,000) and Martinique (64/100,000). Furthermore, SLE prevalence and incidence were found consistently higher in non-white population in the US, Europe, Canada and Australia, which has

been reported in several other studies (Helmick et al., 2008; Johnson et al., 1995; Lau et al., 2006). Underlying factors that may give a plausible explanation for this discrepancy are discussed in section 2.1.6.4.

2.1.6 Etiology

The etiology of SLE is still not completely understood, but multiple factors such as genetic predisposition, environmental exposure, female gender, socioeconomic status (SES), ethnicity and also immunological factors are considered to be important (Molina and Shoenfeld, 2005). Based on these factors, a plausible disease model has been suggested (Arbuckle et al., 2003; Rhodes and Vyse, 2008; Wandstrat and Wakeland, 2001). In this scenario a number of triggers occurring together or sequentially over a limited period of time are required for disease to develop, which happens when a threshold of genetic and environmental susceptibility effects is reached. The genetic background for the individual is determined at birth by inherited susceptibility genes and whether the disease threshold is then reached depends on the environmental influences. For those with many susceptibility genes only a minor environmental trigger may be required; for those with little genetic risk, disease may never develop despite strong or prolonged exposure to the relevant environmental triggers (Figure 1).

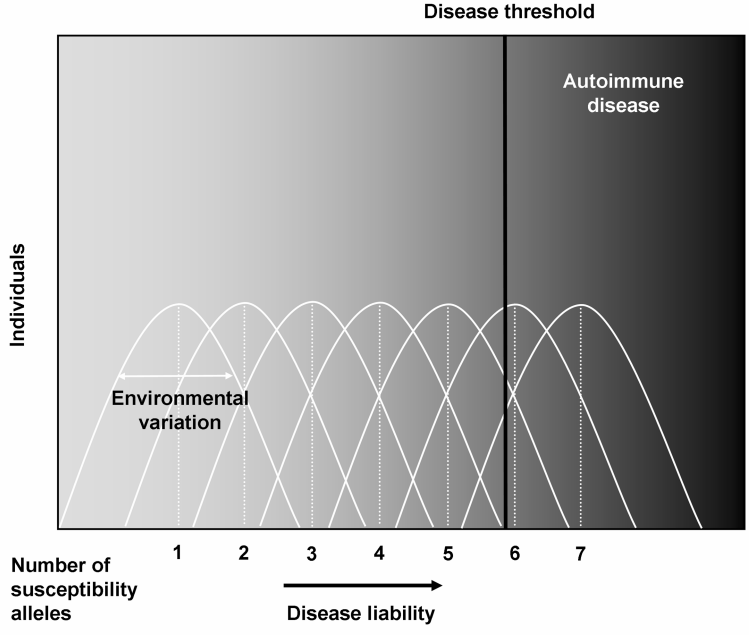

Figure 1. Disease model for autoimmune disease. In this model the x axis represents increasing susceptibility to disease. A gradual increase in the number of susceptibility alleles shifts the disease liability towards the disease threshold and individuals located to the right of this threshold will develop disease. Disease susceptibility is furthermore influenced by environmental and stochastic effects, which is represented by the normal distribution curve. Adapted by permission from Macmillan Publishers Ltd: Nature Immonology (Wandstrat and Wakeland, 2001), copyright 2001.

2.1.6.1 Heredity

Several observations support the importance of a genetic contribution in SLE pathogenesis. For example, familial aggregation studies show a sibling risk ratio (λs) of 20-29 (Alarcon-Segovia et al., 2005; Hochberg, 1987), with 1 being the expected value for diseases lacking familial aggregation. This parameter describes the ratio of the risk to siblings of an affected individual divided by the background population prevalence of the disease (Risch, 1990). A larger value of λs indicates a larger genetic contribution to the disease. Also, family clustering is observed in SLE and 10% to 12% of patients with SLE have a first-degree relative with the disease (Criswell, 2008). Specific traits associated with SLE, such as autoantibody production or complement depletion, have also been observed in the healthy relatives of patients with SLE and may thus be heritable within these families (Rhodes and Vyse, 2008). In addition, first-degree relatives to individuals affected by SLE have an increased risk of autoimmune diseases other than SLE (Alarcon-Segovia et al., 2005; Priori et al., 2003). Twin studies further support a strong genetic component for SLE, with a tenfold concordance ratio of affected monozygotic twins (24-58%) over dizygotic twins (2-5%), when sharing a similar environment (Deapen et al., 1992). However, if SLE was caused by genetics alone one would see a complete concordance in monozygotic twins. This is also illustrated by the fact that the genetic variations associated with SLE are much more common than the prevalence of SLE in the population and thus genetic variation on its own is insufficient to cause SLE per se (Rhodes and Vyse, 2008).

2.1.6.2 Environmental factors

The nature of environmental triggers predisposing to SLE is largely unknown and furthermore results are often conflicting. Irrespective of the environmental trigger responsible, it is most likely a common factor of low penetrance, otherwise we would observe dramatic clustering of cases among individuals with the relevant exposure. Also, it is unlikely to be a single common factor since SLE is not a highly prevalent disease (Rhodes and Vyse, 2008). Several environmental triggers, such as exposure to silica dust, pesticides, certain drugs, hair dyes, high fat/low antioxidant diet, infections, UV, and cigarette smoking have all been associated with SLE, (Molina and Shoenfeld, 2005; Molokhia and McKeigue, 2006; Sarzi-Puttini et al., 2005; Simard and Costenbader, 2007). Some factors, however, have been more consistently associated with the development of SLE. For instance, infection with the Epstein-Barr virus (EBV) appears to be of particular importance (Harley et al., 2006a), with 99% of patients having antibodies against EBV, compared to 90% of the general population (James et al., 2001). Several studies also consistently show increased risk of SLE in smokers, and especially in current smokers. However, when the effect estimates from nine available studies were combined in meta-analysis, only a modestly increased risk was associated with current smoking (RR = 1.50, 95% CI 1.09-2.08) and no increase was seen associated with past smoking (Costenbader and Karlson, 2005).

2.1.6.3 The female sex

There is a marked female predominance in SLE, with a 9:1 female to male ratio, suggesting a role for hormones in SLE susceptibility. In regard to this, oestrogen has been widely studied as a risk factor in SLE, but with conflicting results (reviewed in (Molina and Shoenfeld, 2005; Petri, 2008)). Thus it is likely that a more complex interaction of multiple sex hormones is involved, possibly with a protective effect of male hormones. A potential gene-dose effect of genes located on the X chromosome may also be contributing to the female predominance and this has been suggested for the X-chromosomal SLE susceptibility gene *IRAK1* (Jacob et al., 2007; Jacob et al., 2009). Interestingly, Klinefelter's syndrome (47,XXY) is approximately 14 times higher in men with SLE compared to those without SLE, thus supporting a gene-dose effect from the X-chromosome in SLE susceptibility (Scofield et al., 2008).

2.1.6.4 Ethnic, geographic and socioeconomic factors

The incidence, morbidity and mortality rates are all higher among non-white than white individuals in the United States (reviewed in (Demas and Costenbader, 2009; Lau et al., 2006)). Genetic factors can only partially explain these variations. Low SES, which is a concept that measures income, educational level, wealth, medical insurance, occupation and status or rank in a hierarchical society (Sule and Petri, 2006), as well as a lower sociodemographic position have furthermore been associated with higher incidence, severity and mortality from SLE (Demas and Costenbader, 2009). Race effects may be intricately related to SES and sociodemographic position, which could explain some of the differences seen between different ethnic groups. However, some studies show that African ancestry, independent of all other factors, by itself is associated with higher mortality (Lau et al., 2006). Furthermore comorbidities, such as smoking and exposure to infectious agents, could explain some of the observed disparities related to SES. Consequently this area remains complex and requires further studies.

2.1.6.5 Immunological factors

Dysregulation of immune system also contributes to disease pathology. In some cases this is caused by severe defects in the immune system, such as complement compound deficiencies and impaired phagocytosis, and in others by more subtle defects (reviewed in (Molina and Shoenfeld, 2005)). In this regard it has been shown that in many cases immunological dysfunction, in form of some autoantibodies, precedes the onset of clinical disease (Arbuckle et al., 2003). Antinuclear, anti-Ro, anti-La, and antiphospholipid antibodies usually precede the onset of SLE by many years. Others, including anti-Sm and anti-nuclear ribonucleoprotein antibodies, typically appear only months before diagnosis, during the time when characteristic clinical manifestations appear.

2.1.7 Pathogenesis of SLE

Pathogenic autoantibodies are the primary cause of tissue injury in SLE; however, the detailed pathogenesis leading to the production to these autoantibodies is only partially

understood. Many factors, including dysregulation of T- and B-cells; impaired clearance of apoptotic material and with a possible dysregulation of apoptosis; as well as dysregulation of the expression of certain cytokines, are thought to be the major cause to the development of pathogenic autoantibodies (Figure 2). These factors are in turn a consequence of genetic predisposition, environmental triggers and hormones as previously described. Several of the identified genes predisposing to SLE do indeed have functions related to these pathways (see section 2.3.3). Others are of unknown function or have no apparent role in the processes described here and may thus lead to the identification of new pathways important in SLE pathogenesis.

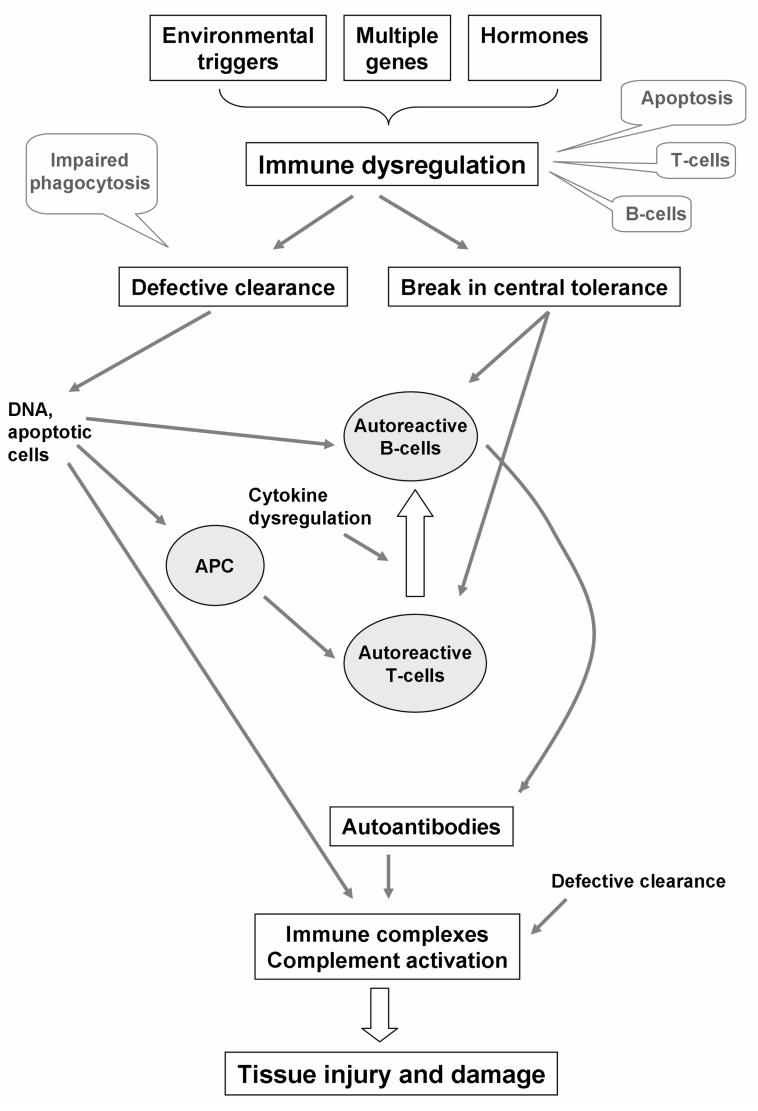

Figure 2. The pathogenesis of systemic lupus erythematosus. Reproduced from [Journal of Clinical Pathology, Mok and Lau, 56: 481-490, 2003] with permission from BMJ Publishing Group Ltd.

2.1.7.1 Autoantibodies

The central immunological disturbance in patients with SLE is the production of pathogenic autoantibodies (Mok and Lau, 2003). These are directed against several self molecules found in the nucleus, cytoplasm, and cell surface, in addition to soluble molecules such as IgG and coagulation factors. As previously described, ANA are the most characteristic for SLE, while anti-dsDNA antibodies are highly specific but less common (section 2.1.3). The importance of antibodies to dsDNA and/or nucleosomes in the pathogenesis of SLE is strongly supported, but the precise mechanism by which their presence actually causes tissue inflammation and damage remains uncertain (reviewed in (Isenberg et al., 2007; Rahman and Isenberg, 2008). The majority of studies of autoantibody-mediated tissue damage in SLE have focused on the kidney, since autoantibodies against dsDNA are commonly found in the kidneys of patients with lupus nephritis. There are two main theories on how these cause tissue damage in patients with SLE, both of which stress that the binding of autoantibodies to dsDNA itself is probably not the most critical determinant of tissue damage (reviewed in (Isenberg et al., 2007; Rahman and Isenberg, 2008). The first model proposes that pathogenic anti-dsDNA autoantibodies bind to nucleosomes in the bloodstream, settle in the renal glomerular basement membrane and subsequently activate complement. The second model proposes a direct pathogenic effect on renal cells through polyreactivity, in which anti-dsDNA autoantibodies, anti-nucleosome autoantibodies, or both, cross-react with proteins in the kidney and activate complement. The pathogenesis of manifestations other than glomerulonephritis is less well understood, although immune complex deposition with activation of complement at relevant sites is a probable mechanism (Mok and Lau, 2003).

2.1.7.2 Apoptosis and clearance

Apoptosis and clearance of apoptic cells/material are considered key processes in the etiology of SLE, and deficiencies in the recognition and clearance of apoptotic cells by phagocytosis have been shown in patients with SLE (reviewed in (Janko et al., 2008; Mok and Lau, 2003; Munoz et al., 2008)). Whether apoptosis itself is abnormal or merely an effect of environmental triggers, such a UV irradiation or viral infection, is less understood (reviewed in (Cohen, 2006; Harley et al., 2009)). During necrosis or apoptosis nuclear antigens will be subjected to modifications which will be recognized as non-self. Normally, phagocytes will quickly remove apoptotic cells and blebs long before they could have released their modified contents. However, if the apoptotic cells are not removed effectively they will start to spill out these modified autoantigens, which will subsequently be presented to antigen-presenting cells (APC) and trigger autoimmunity.

2.1.7.3 Dysregulation of T- and B-cells

Pathogenic autoantibodies are produced by B-cells in the presence of stimulating antigen. In general, this process can occur only in B-cells that are being co-stimulated

by T-cells. In healthy individuals the presence of foreign antigens such as bacteria and viruses is required and B-cells that have the ability to interact with self-antigens are either removed, made inactive or have their antibodies edited so that they can not bind antigen. However, in patients with SLE both B- and T-cells specific to self-antigens are allowed to remain viable and by the interaction of these cells the production of high affinity IgG autoantibodies is made possible (reviewed in (Rahman and Isenberg, 2008)).

2.1.7.4 Cytokines

Cytokine profiles in patients with SLE have been studied extensively and subsequently several cytokines have been implicated in SLE pathogenesis. Pro-inflammatory cytokines in particular play an important role in propagating the inflammatory process responsible for tissue damage. Some pro-inflammatory cytokines are found over-expressed in patients with SLE and/or correlate with disease severity, including interleukin (IL)-6 (Grondal et al., 2000; Linker-Israeli et al., 1991), tumour necrosis factor (TNF)-α (reviewed in (Aringer and Smolen, 2008)), interferon (IFN)-γ (Viallard et al., 1999), IL-18 and IL-12 (Wong et al., 2000). IL-10 is also elevated in patients with active SLE and correlates with disease activity and even though this cytokine is classically considered as anti-inflammatory it appears to have an inflammatory role in SLE (reviewed in (Ramanujam and Davidson, 2008)). A dual role in inflammation appears to be the case also for transforming growth factor (TGF)-β, which has both anti-inflammatory functions and a role in promoting inflammatory T helper (Th) 17-cells, which have been associated with autoimmune inflammation (reviewed in (Diveu et al., 2008)). Th17-cells produce IL-17, which in turn is dependent on IL-23. Both these cytokines have been found elevated in SLE patients (Wong et al., 2008). Furthermore, patients with SLE have increased serum levels of IFN-α, which correlate to both disease activity and severity and increased expression of IFN-α-regulated genes (reviewed in (Crow and Kirou, 2004; Ronnblom and Pascual, 2008)). Interestingly, several genes in this pathway have been found associated to SLE (see 2.3, Genetics of SLE).

2.2 GENETIC MAPPING IN COMPLEX DISEASES

Genetic mapping in complex diseases, defined as the localization of genes underlying phenotypes on the basis of correlation with DNA variation, is indeed complex based on a number of reasons (reviewed in (Altshuler et al., 2008; Criswell, 2008). First of all, complex diseases are not inherited in a straightforward Mendelian fashion (i.e. monogenic inheritance). Second, the inherited genotype does not always correspond to the resultant phenotype (explained by incomplete penetrance of predisposing loci or disease that develops in the absence of apparent genetic risk factors). Third, there are at least several disease predisposing genes, all with modest effects (odds ratios (ORs) are usually < 1.5). Forth, not all affected individuals for a particular disease share the same genetic risk (genetic heterogeneity), and finally, there is an important role for environmental factors.

Complex diseases are relatively common in the population. In general, complex diseases have a late onset and therefore their impact on reproductive fitness is modest or absent, which allows causative alleles to rise to moderate frequencies in the population. Also, it is thought that some alleles that were advantageous or neutral during human evolution, might now confer susceptibility to disease because of changes in living conditions accompanying civilization. Disease-causing variants may also be maintained at high frequency when the disease burden is counterbalanced by a beneficial phenotype. In line with this, the "common disease–common variant" hypothesis was suggested, which proposes that common polymorphisms (variants with a minor allele frequency [MAF] > 1%) might contribute to susceptibility to common diseases (reviewed in (Altshuler et al., 2008)). This, however, does not mean that all causal mutations are common, only that some common variants exist and could be used to pinpoint loci for detailed study.

Before 2006 the traditional manner in which susceptibility genes were identified was either through positional mapping approaches or by functional candidate gene studies. Positional mapping approaches are hypothesis-free and consist of two steps: an initial genome-wide linkage analysis followed by refinement of the identified chromosomal region by association analysis. The candidate gene association study is on the other hand based on the hypothesis that certain genes influence disease susceptibility and the genetic markers that are tested are based on this assumption. Genetic linkage and association analysis basically rely on similar principles, i.e. the co-inheritance of adjacent DNA variants. Linkage relies on identifying haplotypes that are inherited intact over several generations and association relies on the retention of adjacent DNA variants over many generations (Cardon and Bell, 2001). Thus, association studies can be regarded as very large linkage studies of unobserved, hypothetical pedigrees. As a result of the rapid technological advances in genotyping methods it is today possible to genotype a million single nucleotide polymorphisms (SNPs) in one individual at a time, which constitutes the basis of a genome-wide association (GWA) study. The GWA analysis combines the power and resolution of a conventional association study with the hypothesis-free methodology of a genome-wide linkage scan. In 2006 eight GWA

studies were published and by the beginning of 2009 the number had grown to 398 (www.genome.gov/gwastudies) and GWA studies are now the most widely used approach for genetic mapping (McCarthy et al., 2008).

2.2.1 Sequence variation in the human genome

Genetic variations in the human genome can be either common or rare. Common variations are defined as genetic variants with a MAF of at least one percent in the population and are in general synonymous with polymorphisms, while variations occurring in less than one percent are generally defined as mutations (reviewed in (Frazer et al., 2009)). Genetic variation includes SNPs, simple sequence repeats (SSRs, also known as micro- and minisatellites), as well as structural variants, which include insertions and deletions (indels), block substitutions, inversions, copy number variants (CNVs), segmental duplication and translocations (Figure 3) (Feuk et al., 2006). The vast majority of genetic variants are hypothesized as being neutral, i.e. they do not contribute to phenotypic variation, but the relative percentage of neutral and non-neutral variants is not yet clear (reviewed in (Frazer et al., 2009)).

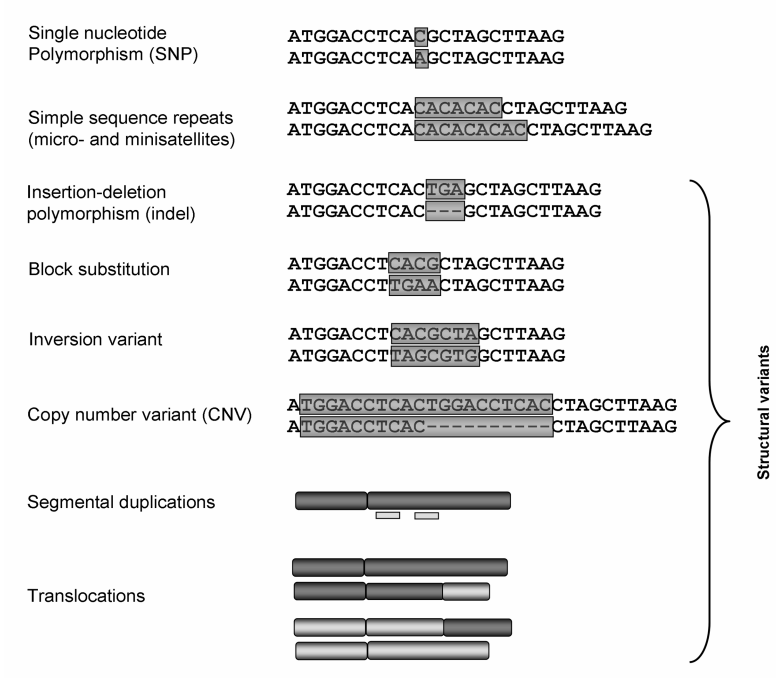

Figure 3. Classes of human genetic variation. DNA sequence variations affecting a single nucleotide are known as single nucleotide polymorphisms. Indels occur when base pairs are present in some genomes but absent in others. SSRs are short tandem repeat units, block substitutions are string of adjacent nucleotides that varies between two genomes, while in an inversion variant the order of the base pairs is reversed in a defined section of a chromosome. CNVs occur when identical or nearly identical sequences are repeated in some chromosomes, while segmental duplications are repeated segments with near-identical sequence. Translocations are rearrangements of chromosomal sections between non-homologous chromosomes. Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Genetics (Frazer et al., 2009), copyright 2009.

2.2.1.1 Single nucleotide polymorphisms

SNPs constitute the majority of genetic variation in the human genome and comprise single base-substitutions or single base insertion/deletion. The number of SNPs in any Caucasian genome is approximately 3.3 million, with an average of 1 SNP in 1000 bases (reviewed in (Altshuler et al., 2008; Frazer et al., 2009). Approximately 2% of SNPs are estimated to be of biological importance (reviewed in (Orr and Chanock, 2008)). Non-synonymous SNPs are located in the protein-coding regions and cause amino acid substitutions, frame shifts or termination of protein translation. Synonymous SNPs, i.e. SNPs located within exons but that does not alter protein primary structure, have been shown to affect mRNA stability or alter splicing signals. SNPs located outside of the protein coding regions can also be of functional importance, either by their location in gene promoters where they may affect gene regulation by altering transcription binding sites, or by their location in enhancers or silencers.

2.2.1.2 Simple sequence repeats

SSRs comprise about 3% of the human genome and are short tandem repeat units composed of either 1±13 bases, often termed microsatellites, or 14±500 bases, often termed minisatellites, with approximately 1 SSR per 2 kb (Lander et al., 2001). SSR most commonly consist of di-, tri- and tetranucleotide repeats, and show a high degree of heterogeneity within a population. Subsequently SSRs, and microsatellites in particular, have been very useful as genetic markers in the mapping of human disease (see 2.2.2, linkage analysis) and to establish relatedness between individuals. Large expansions of trinucleotide repeats can lead to genomic instability, however, the impact of SSRs of modest length on disease remains to be determined (reviewed in (Orr and Chanock, 2008)).

2.2.1.3 Structural variants

There is no common consensus on how to define structural variations in the human genome. In the purpose of simplicity, structural variations are here defined as all base pair variations that are not SNPs or SSRs (Figure 3). It has been estimated that structural variations constitutes between 9 and 25 Mb of an individual human genome and thus structural variants are likely to make an important contribution to human diversity and disease susceptibility (reviewed in (Feuk et al., 2006; Frazer et al., 2009). Structural variants appear to have a similar behavior to SNPs in regards of both genomic and population distribution, indicating a similar evolutionary history (reviewed in (Frazer et al., 2009)). Several studies have shown that common short indels, CNVs, as well as larger common structural polymorphisms in unique regions of the genome are in linkage disequilibrium (LD) with tagging SNPs (reviewed in (Frazer et al., 2009; Wain et al., 2009)), and thus nearby SNPs can serve as proxies for common structural variants in association analyses (see 2.2.3, association analysis).

2.2.2 Linkage analysis

Linkage analysis is used to trace co-segregation of a disease gene with a genetic marker, in the past often a microsatellite (see 2.2.1.2), within families where the disease is inherited over several generations (reviewed in (Borecki and Province, 2008)). Two genetic loci are linked when they are sufficiently close together on a chromosome, so that their alleles tend to co-segregate within families. However, co-segregating loci may be separated by the process of recombination, which is less likely to occur when two loci are close together and more likely between loci that are located far apart. The probability of recombination is a measure of the genetic distance between two loci, where two loci that show 1% recombination is defined as being 1 centimorgan (cM) apart (approximately 1 Mb of DNA). Two loci located far apart on the same chromosome or on two different chromosomes will segregate independently, meaning that on average 50% of gametes will be recombinant and 50% will be non-recombinant. In a linkage analysis the aim is to identify loci for which the probability of recombination is less than 50%, by estimating the recombination between a disease locus and individual markers with known position. This is done by calculating the probability of linkage versus no linkage within a pedigree. The ratio of these two likelihoods gives the odds of linkage, which is reported as the logarithm of the odds (LOD) score (defined by (Morton, 1955)). Higher values of the LOD scores support the hypothesis of linkage, while negative values of the LOD score gives evidence for independent assortment.

When the mode of inheritance is known, i.e. in Mendelian disorders, standard parametric linkage analysis can be used, consequently because of the inheritance model, the gene frequency and the penetrance of each genotype must be specified. In complex diseases this is rarely the case and therefore a model-free, or non-parametric, method is used for linkage analysis. Instead of testing whether the inheritance pattern fits a specific model for a trait-causing gene, which is done in standard parametric methods, non-parametric methods test whether the inheritance pattern deviates from what is expected for independent assortment (Kruglyak et al., 1996). Linkage analysis has been tremendously successful for mapping genes in monogenic Mendelian diseases, in which the causing variants are often rare and of high penetrance, but has less power for detecting common alleles that have low penetrance and modest effects on disease, which is often the case in complex diseases (reviewed in (Hirschhorn and Daly, 2005)). Because linkage focus only on recent ancestry, with the opportunity of only a few recombinations to occur, the identified regions will usually span tens of Mb and encompass several hundred of potential candidate genes (Cardon and Bell, 2001).

2.2.3 Association analysis

Association analysis is more powerful than a linkage study to detect genetic contributions to complex disease (LD) (Risch and Merikangas, 1996; Risch, 2000) and is based on the comparison of differences in allele frequencies between cases and appropriate controls. The control individuals can either be unrelated, matched individuals or unaffected family members. In contrast to a linkage analysis, association

looks for historical recombination within populations across hundreds or thousands of generations. When a disease-causing mutation arises on a particular copy of the human genome it will be co-inherited with a sequential set of nearby located markers. Since the probability of recombination is low between adjacent markers, disease alleles in the population typically show association with nearby marker alleles for many generations (reviewed in (Altshuler et al., 2008; Borecki and Province, 2008)). This correlation between nearby variants is known as linkage disequilibrium (LD). In addition to de novo mutations, genetic drift, admixture of populations with distinct evolutionary histories, and rapid population expansion also influence LD (reviewed in (Borecki and Province, 2008)). Segments of the genome that show strong LD, i.e. spanning markers that are strongly correlated with each other, are referred to as haplotype blocks (Daly et al., 2001). By characterizing SNP frequencies and local LD patterns across the human genomes with Asian, African and European ancestry, the International HapMap project (www.hapmap.org) have shown that that the vast majority of common SNPs are strongly correlated to one or more nearby proxies and that the LD patterns are remarkably stable over different samples of individuals (reviewed in (Altshuler et al., 2008; Borecki and Province, 2008)). Within haplotype blocks, it is possible to infer genotypes of common SNPs based on the knowledge of only a few empirically determined tag SNPs and it has been shown that approximately 500,000 SNPs provide excellent power to test >90% of common SNP variation in non-African populations (reviewed in (Altshuler et al., 2008)). Haplotype blocks in individuals of African descent tend to be smaller and greater in number than those in European/Caucasian populations, consistent with this population being older and thus have undergone more recombination, and requires almost double the amount of SNPs to obtain similar power (Altshuler et al., 2008).

The most common strategy for gene identification by association is the case-control study design. This is based on a comparison between cases, ascertained for a specific phenotype and therefore assumed to have a high prevalence of susceptibility alleles, and controls, not ascertained for the phenotype and considered likely to have a lower prevalence of such alleles (McCarthy et al., 2008). The advantage with a case-control study design is that the study material is more easily obtained than families and that this approach is potentially a powerful strategy for identifying genes of small effects that contribute to complex traits (Cardon and Bell, 2001; Risch, 2000). However, the case-control study design may be sensitive to population stratification; i.e. the presence of individuals with different ancestral and demographic histories and therefore different allele frequencies independent of disease, which can lead to spurious associations (Lander and Schork, 1994). Furthermore, so-called ‘‘cryptic relatedness’’, which is defined as family relationships among the cases or controls that is not known to the investigator, might also potentially lead to false positive associations (Voight and Pritchard, 2005).

To overcome the problem with population stratification, family-based association studies can be used, basically because these methods use the untransmitted parental alleles as controls. The transmission disequilibrium test (TDT) focuses on specific

alleles transmitted to affected offspring from heterozygous parents in parent-affected offspring trios, thereby providing a joint test of linkage and association (Spielman et al., 1993). In principle, this test is similar to that of a case-control analysis, but differs in how the number expected is computed under the null hypothesis (reviewed in (Borecki and Province, 2008; Laird and Lange, 2008)). In a case-control study, an equal distribution of alleles in both cases and controls is expected under the null hypothesis. In TDT, the null hypothesis is based on the rules of Mendelian inheritance, i.e. that the putative disease-associated allele is transmitted 50% of the time from heterozygous parents, with the alternative hypothesis being that the disease-associated allele will be transmitted more often to affected offspring. In its original form the TDT requires genotypes from both biological parents, of which at least one parent must be heterozygous to be informative, as well as from the affected offspring. Consequently transmissions from homozygous parents are not used and thus the effective sample size may be considerably less than the number of trios, depending on allele frequency, which reduces the statistical power to detect association when compared with a case-control analysis (reviewed in (Cardon and Bell, 2001; Laird and Lange, 2008)). There are several extensions of the TDT dealing with missing parental genotypes, for instance using phenotypically discordant sib pairs (reviewed in (Borecki and Province, 2008)). Often data are available for larger pedigrees with multiple nuclear families and/or discordant sibships and for this purpose the pedigree disequilibrium test (PDT) for analysis of LD in general pedigrees was developed (Martin et al., 2000). This test uses data from related nuclear families and discordant sib pairs from extended pedigrees. Furthermore, the test retains a key property of the TDT, in that it is valid even when there is population substructure. Power simulations demonstrate that, when extended pedigree data are available, substantial gains in power can be attained by use of the PDT.

2.2.3.1 The positional mapping approach

In positional mapping, an association analysis is performed in a region previously shown to be linked to disease (reviewed in (Wang et al., 2003)). With this approach the initial linkage studies provide information on both the position and the genetic effect of underlying disease loci, while the association mapping extends linkage analysis to map the position of the disease gene to a much higher resolution. Positional mapping has been successful in rare Mendelian disorders and has suggested novel disease mechanisms also in complex diseases (reviewed in (Wang et al., 2003)). However, since linkage regions are large in complex disorders, the identification of disease genes is cumbersome using this approach, with several steps of association mapping required.

2.2.3.2 The candidate gene approach

This approach tests specific candidate genes for association, based on existing knowledge of the disease pathogenesis, functions of the selected genes, and in some cases, data from animal models of the disease (Wang et al., 2003). In addition, location within a previously linked region could guide the selection of functional candidate genes. A central problem with this approach is that it is usually not very

straightforward, given that the knowledge of the underlying disease pathogenesis is limited and thus each candidate gene has a tiny a priori probability of being disease-causing. Although many claims of associations have been published for complex diseases, the statistical support tends to be weak and few of the published gene findings have been replicated (Hirschhorn et al., 2002; Lohmueller et al., 2003).

2.2.3.3 The genome-wide association study approach

The GWA approach is based on association but does not rely on any prior hypothesis regarding position or function and aims to cover the majority of the common variants in the genome by using knowledge of LD relationships. Even though GWA studies have been very successful in identifying susceptibility genes there are still limitations (Altshuler et al., 2008; Frazer et al., 2009). First, GWA studies generally identify only common genetic variants and the studies performed so far have had good power to detect alleles that are common in the general population and have modest effect sizes. However, as many as 8,600 samples are required to provide sufficient power for detection of an allele with a frequency of 20% and an OR of 1.2, (Altshuler et al., 2008). Thus, rare variants or those with low effect sizes are likely to have been missed in current GWA study designs. Second, the coverage of GWA scans are non-random, meaning that subsets of genomic regions are poorly covered (Frazer et al., 2009). Thirdly, in most cases the association signals identified in GWA studies are likely to be indirect associations due to LD and not the causal variants themselves (Altshuler et al., 2008; Frazer et al., 2009).

2.2.4 Identification of causal variants

The basic idea behind genetic mapping is not primarily risk prediction, but rather to understand the mechanisms underlying a specific disease. Thus, when a gene is identified it has yet to be scrutinized by fine mapping and resequencing for identification of causal variants. In those cases where the causative variants are located in exons and truncate or otherwise alter the gene product the identification is usually straightforward. However, in complex disease the causal variants are more often non-coding and rather more likely to have regulatory roles or to be structural variants (Frazer et al., 2009). Interestingly, roles for structural variants in complex traits have recently been shown in both autism and schizophrenia (reviewed in (Frazer et al., 2009)). Regulatory variants can be located in promoters, introns, or even in more distant sites located hundreds of kb from nearby genes. Furthermore, comparative genome analysis has suggested that 5% of the human genome is functional and less than one-third of this consists of genes that encode proteins (Waterston et al., 2002), thus roles for causative non-coding variants are likely. The roles of regulatory variants are several and besides a role in transcription they could affect the stability, localization and translation of messenger RNA (mRNA). Thus, to fully understand the role of regulatory variants in disease processes functional studies are of vital importance. Comparing differences in expression, distribution and splice variants of mRNA and protein between patients and controls, as well as identifying differences in the binding of transcription factors could be useful when studying regulatory variants. Given this

complexity the path from gene identification to fully understand the role of this gene in disease is usually not straightforward and for many susceptibility genes their role in diseases still remains to be elucidated.

2.2.5 Gene-gene interaction (epistasis)

Gene-gene interaction, or epistasis, is a phenomenon that occurs when the effect of one locus is being altered by effects at another locus, i.e. the effect of carrying more than one variant is different than would be expected by simply combining the effects of each individual variant (reviewed in (Cordell, 2002; Hirschhorn and Daly, 2005; Moore and Williams, 2005)). It is important to keep in mind that statistical tests of interaction are limited to testing specific hypotheses based on mathematical models, where a departure from any of these models is defined as epistasis (reviewed in (Cordell, 2002)). Thus, statistical interaction does not necessarily imply interaction on the biological level and thus may not easily translate to physical interactions between proteins.

2.2.6 The future of genetic mapping in complex disease

The introduction of GWA studies in genetic mapping of complex diseases have rapidly increased the number of candidate genes associated with diseases. However, GWA studies to date have had low statistical power to capture common variants of lower frequency (0.5 to 5%) as well as to identify gene-gene interactions (Altshuler et al., 2008; Hirschhorn and Daly, 2005). Thus many more disease loci remain to be identified, something that will be facilitated by meta-analyses, recruitment of larger sample sets, and inclusion of samples of non-European ancestry (reviewed in (Altshuler et al., 2008). To successfully identify structural variants and low frequency common variants, a comprehensive catalog of genomic variation as well as characterization of LD relationships will be required, which will most likely be achievable with new massive parallel sequencing technologies (reviewed in (Altshuler et al., 2008; Frazer et al., 2009)). In addition to the genetic inheritance exposures to, and interaction with, environmental factors likely play a large role in human phenotypic variation. However, these are more difficult to measure and thus improved methods will be required to fully understand the role of environmental exposure in complex diseases (Altshuler et al., 2008). Thus, given all the complexity involved in the genetic mapping of complex diseases it is unlikely that 100% of the genetic variation will be explained in the years to come (Altshuler et al., 2008).

2.3 GENETICS OF SLE

The genetic contribution to SLE has been known for at least 30 years, shown both by familial aggregation and twin studies (see section 2.1.6.1), yet the identification of susceptibility genes underlying SLE pathogenesis has been sparse until recently. Before 2007, nine established SLE susceptibility genes had been identified through linkage approaches or candidate gene association studies (Harley et al., 2009; Moser et al., 2009). Today, as a result of increasingly powerful genetic approaches, and particularly GWA studies, the number of convincingly identified genes is now well over 30 (Moser et al., 2009). This progress, with the major genetic discoveries and the parallel identification of SLE susceptibility genes is illustrated in Figure 4, while Table 2 lists the genes that have shown convincing association with SLE in one or several studies. Genetic analyses have the unique capability of identifying causative disease mechanisms and given that they are usually performed in hypothesis-free manner they have the potential to reveal novel and fundamental keys to understanding SLE pathogenesis.

One important aspect to consider, when studying the genetic contribution in SLE, is its heterogeneity, given that only 4 out of a possible 11 ACR criteria (Tan et al., 1982) are required to make a diagnosis of SLE (Hochberg, 1997). Thus it is possible that not all patients with SLE have the exact same disease and that different SLE manifestations have different genetic contribution. One strategy to eliminate heterogeneity is to stratify the sample set according to sub-phenotypes, for example lupus nephritis, however, with the cost of reduced power to detect association given the smaller sample set.

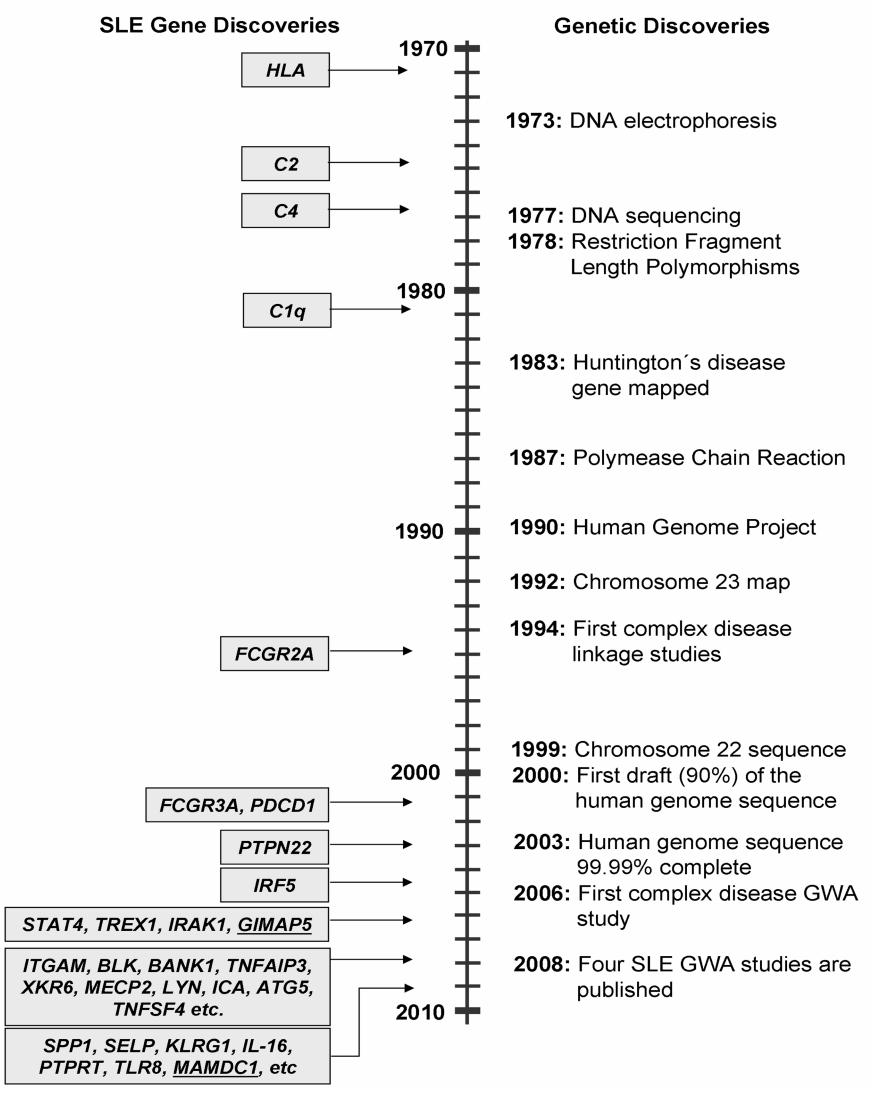

Figure 4. A timeline of discoveries in human genetics (right) and genes discovered in SLE (left). Underlined genes have been identified as part of this thesis. Adapted by permission from Macmillan Publishers Ltd: Genes and Immunity (Moser et al., 2009), copyright 2009.

Table 2. SLE candidate genes and their ways of identification.

Genes/loci	Ways of identification	References
FCGR2A	Identified by positional mapping.	(Harley et al., 2008; Karassa et al., 2002)
MHC	Confirmed in GWA studies.	(Graham et al., 2008; Harley et al., 2008; Hom et al., 2008)
IRF5-TNPO3	Identified in association studies. Confirmed in GWA studies.	(Cunninghame Graham et al., 2007; Graham et al., 2008; Harley et al., 2008; Hom et al., 2008; Kyogoku and Tsuchiya, 2007; Sigurdsson et al., 2005)
STAT4		(Harley et al., 2008; Hom et al., 2008; Remmers et al., 2007; Sigurdsson et al., 2008; Taylor et al., 2008)
PTPN22		(Harley et al., 2008; Kyogoku et al., 2004)
ITGAM-ITGAX	Simultaneously identified by positional mapping and by GWA studies.	(Harley et al., 2008; Hom et al., 2008; Nath et al., 2008)
FAM167A-BLK	Identified in GWA studies.	(Graham et al., 2008; Harley et al., 2008; Hom et al., 2008)
TNFAIP3		(Graham et al., 2008; Musone et al., 2008)
PXK		(Harley et al., 2008)
BANK1		(Kozyrev et al., 2008)
ATG5		(Harley et al., 2008)
ICA1		(Harley et al., 2008)
XKR6		(Harley et al., 2008)
LYN		(Harley et al., 2008)
KIAA1542 (IRF7)		(Harley et al., 2008)
Nogene (1q25.1)		(Harley et al., 2008)
SCUBE1		(Harley et al., 2008)
NMNAT2		(Harley et al., 2008)
UBE2L3		(Harley et al., 2008)
C8orf12		(Harley et al., 2008)
FCGR3A	Identified by positional mapping.	(Aitman et al., 2006; Edberg et al., 2002; Fanciulli et al., 2007)
CRP		(Edberg et al., 2008; Russell et al., 2004)
PDCD1		(Prokunina et al., 2002)
MAMDC1		(Hellquist et al., 2009)
TYK2	Identified in association studies.	(Graham et al., 2007; Sigurdsson et al., 2005)
MECP2		(Sawalha et al., 2008; Webb et al., 2009)
SPP1		(Han et al., 2008)
IRAK1		(Jacob et al., 2007; Jacob et al., 2009)
TREX1		(Lee-Kirsch et al., 2007)
C1Q		(Walport et al., 1998)
C4A		(Fielder et al., 1983; Yang et al., 2007)
C4B		(Fielder et al., 1983; Yang et al., 2007)
C2		(Sullivan et al., 1994)
TNFSF4		(Cunninghame Graham et al., 2008)
SELP		(Jacob et al., 2007; Morris et al., 2009)
CTLA4-ICOS		(Graham et al., 2006; Lee et al., 2005)
KLRG1		(Armstrong et al., 2009)
IL-16		(Armstrong et al., 2009)
PTPRT		(Armstrong et al., 2009)
TLR8		(Armstrong et al., 2009)
MBL		(Garred et al., 2001)
GIMAP5		(Hellquist et al., 2007; Lim et al., 2009)

Legend for Table 2: *ATG5*; autophagy related 5 homolog; *BANK1*, B-cell scaffold protein with ankyrin repeats 1; *BLK*, B lymphoid tyrosine kinase; *C1Q*, complement component 1, q subcomponent; *C2*, complement component 2; *C4A*, complement component 4A; *C4B*, complement component 4B; *C8orf12*, chromosome 8 open reading frame 12; *CRP*, C-reactive protein; *CTLA4*, Cytotoxic T-cell Antigen 4; *FAM167A*, family with sequence similarity 167, member A; *FcGR2A*, Fc fragment of IgG, low affinity IIa, receptor (CD32); *FcGR3A*, Fc fragment of IgG, low affinity IIIa, receptor (CD16a); *GIMAP5*, GTPase of the immune-associated protein 5; *ICA1*, islet cell autoantigen 1, 69kDa; *IL-16*, interleukin 16; *ICOS*, Inducible T-cell costimulator; *IRAK1*, interleukin-1 receptor associated kinase-1; *IRF5*, interferon regulatory factor 5; *IRF7*, interferon regulatory factor 7; *ITGAM*, integrin, alpha M; *ITGAX*, integrin alpha X; *KIAA1542*, PHD and ring finger domains 1; *KLRG1*, killer cell lectin-like receptor 1; *LRFN5*, leucine rich repeat and fibronectin type III domain containing 5; *LYN*, v-yes-1 Yamaguchi sarcoma viral related oncogene homolog; *MAMDC1*, MAM domain containing 1; *MBL*, mannose binding lectin; *MECP2*, methyl CpG binding protein 2; *MHC*, major histocompatibility complex; *NMNAT2*, nicotinamide nucleotide adenylyltransferase 2; *PDCD1*, programmed cell death 1; *PTPN22*, protein tyrosine phosphatase, non-receptor type 22; *PTPRT*, protein tyrosine phosphatase receptor type T; *PXK*, PX domain containing serine/threonine kinase; *SCUBE1*, signal peptide, CUB domain, EGF-like 1; *SELP*, P-selectin; *SPP1*, osteopontin; *STAT4*, signal transducer and activator of transcription 4; *TLR8*, toll-like receptor 8; *TNFAIP3*, tumour necrosis factor-a induced protein 3; *TNFSF4*, tumour necrosis factor receptor superfamily, member 4; *TNPO3*, transportin 3; *TREX1*, three prime repair exonucleas; *TYK2*, Tyrosine kinase 2; *UBE2L3*, ubiquitin-conjugating enzyme E2L 3; *XKR6*, XK, Kell blood group complex subunit-related family, member 6. Adapted from (Harley et al., 2009; Moser et al., 2009; Rhodes and Vyse, 2008). Bold genes have been identified through studies included in the present thesis.

2.3.1 Linkage studies in SLE

Linkage studies, although very successful in identifying causative loci in single gene disorders, have not generated a large number of candidate genes in SLE even though a substantial number of susceptibility regions have been identified (reviewed in (Harley et al., 2006b; Rhodes and Vyse, 2007). As reviewed in section 2.2.2, linkage studies search for genomic loci that are co-transmitted with the disease in families containing two or more affected members (Lander and Schork, 1994). As a result of the genome-wide scans performed in SLE (Cantor et al., 2004; Gaffney et al., 1998; Gaffney et al., 2000; Gray-McGuire et al., 2000; Johansson et al., 2004; Koskenmies et al., 2004a; Lindqvist et al., 2000; Moser et al., 1998; Nath et al., 2004b; Olson et al., 2002; Shai et al., 1999; Xing et al., 2007), ten regions spanning chromosomes 1q22-24, 1q31-32, 1q41-42, 2q37, 4p16, 5p15-5q11, 6p11-21, 11p13, 12q24 and 16q12-13 and showing strong linkage with independent confirmation have been identified (reviewed in (Harley et al., 2006b; Rhodes and Vyse, 2007). None of these regions have, however, been reproduced in all studies. However, the cohorts that have been used in the major linkage studies have included varying proportions of different ethnicities and thus genetic heterogeneity may partly explain the lack of consistency.

Susceptibility genes showing convincing association have been identified in three of these regions; *FCGR2A*, *FCGR2B*, *FCGR3A*, *CRP* on 1q22-24, *PDCD1* on 2q37 and the major histocompatibility complex (*MHC*) on 6p11-21, illustrating the difficulties involved in identifying susceptibility genes within a linkage region. Fcγ-receptors mediate clearance of immune complexes, while CRP is an acute-phase protein that interacts with Fcγ-receptors on the surface of leukocytes and induce phagocytosis and release of inflammatory cytokines through interaction with FcγR2A (reviewed in (Rhodes and Vyse, 2007)). SNPs at each of the low-affinity Fcγ-receptor genes have been associated with SLE, although early reports are from smaller studies and results

are often inconsistent (Brown et al., 2007). However, recent studies show strong support for *FCGR2A* through a meta-analysis as well as one of the GWA studies (Harley et al., 2008; Karassa et al., 2002). *PDCD1* encodes for a cell surface receptor of the CD28 family that is expressed on the surface of B- and T-cells and mediates down-regulatory signals that inhibit T-cell proliferation and cytokine secretion. The A allele of rs11568821 (PD1.3) in *PDCD1*, suggested to alter the binding site for the runt-related transcription factor 1 (RUNX1), has been found associated with SLE (Prokunina et al., 2002). Genes within the *MHC* have shown strong and consistent linkage with SLE in a large number of studies, with additional confirmation in GWA studies (reviewed in (Criswell, 2008; Harley et al., 2009; Moser et al., 2009; Rhodes and Vyse, 2007, 2008; Tsao, 2004; Wakeland et al., 2001)). The *MHC* contains at least 250 expressed genes of which several are inherited together as one block due to the high degree of LD in this region (Rhodes and Vyse, 2008). As a consequence it is difficult to establish whether an associated variant is functional or in LD with functional polymorphisms elsewhere. Recent work has, however, shown evidence of two distinct and independent association signals arising from both *HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1* within the Class II region and the *SKIV2L* (superkiller viralicidic activity 2-like) within the Class III region (reviewed in (Criswell, 2008; Rhodes and Vyse, 2008)). Furthermore, one of the studies included in this thesis describes the identification of the SLE candidate gene *MAMDC1*, which was found through fine mapping of a previously identified linkage region on chromosome 14q21-q23 (Paper II) (Koskenmies et al., 2004a).

Two genome-wide linkage scan meta-analyses have further identified two significantly linked regions; 16p12.3–16q12.2 (Lee and Nath, 2005) and 20p11–q13.13 (Forabosco et al., 2006). A fine mapping of the 16p12.3–16q12.2 region recently identified association to the *ITGAM* gene, also known as CD11b or complement receptor 3 (CR3), which is expressed primarily on neutrophils, macrophages and dendritic cells and can bind a variety of ligands including intercellular adhesion molecule 1 (ICAM-1) (Nath et al., 2008). Simultaneously this gene was shown to be strongly associated to SLE in two of the GWA studies (Harley et al., 2008; Hom et al., 2008).

Notably, convincing SLE candidate genes such as *IRF5* and *BLK*, which have been identified through association studies and furthermore confirmed in GWA studies, are not located in any of the regions that have been found significantly linked to SLE. Thus, it is apparent that linkage studies have limitations in identifying susceptibility genes for complex diseases such as SLE. As described in section 2.2.2, linkage studies have less power for detecting common alleles that have low penetrance and modest effects on disease, which is likely to be the case for many of the disease causing variants in SLE.

2.3.2 Association studies in SLE

The vast majority of SLE candidate genes have been identified through association studies, which either constitutes candidate gene association studies or GWA studies (Table 2). As previously described in section 2.2.3, association studies test for a difference in SNP frequency between cases and controls.

Previous to the introduction of the GWA approach, most association studies were performed using a rather small number of SNPs per project due to the limitations in technology and also the heavy costs of genotyping. As a consequence most studies were restricted to a single gene or gene clusters, which were selected based on their biological function and/or position within a linkage region. In a GWA study, however, hundreds of thousands of SNPs can be simultaneously genotyped across the whole genome. A GWA study is similar to that of a genome-wide linkage scan, given that both methods are based on a hypothesis-free methodology. Yet, the power and resolution of a GWA study is higher than of a linkage scan. Three high-density GWA studies (Graham et al., 2008; Harley et al., 2008; Hom et al., 2008) and one smaller study (Kozyrev et al., 2008) as well as several candidate gene studies (reviewed in (Criswell, 2008; Harley et al., 2009; Moser et al., 2009; Rhodes and Vyse, 2008)) have been performed up to date. These have identified the majority of SLE susceptibility genes, some with very high levels of evidence of association and some requiring further studies due to either low levels of significance or inconsistencies between studies (differences in effects or in the key SNPs showing association). Thus, association studies have proven to be a powerful tool in the identification of SLE susceptibility genes.

2.3.3 SLE susceptibility genes and their roles in pathogenesis

Many of the SLE susceptibility genes (Table 2) have functions related to the pathways which have been previously shown to be dysregulated or impaired in patients with SLE (section 2.1.7) and which are described in the following sections, as well as graphically in Figure 5.

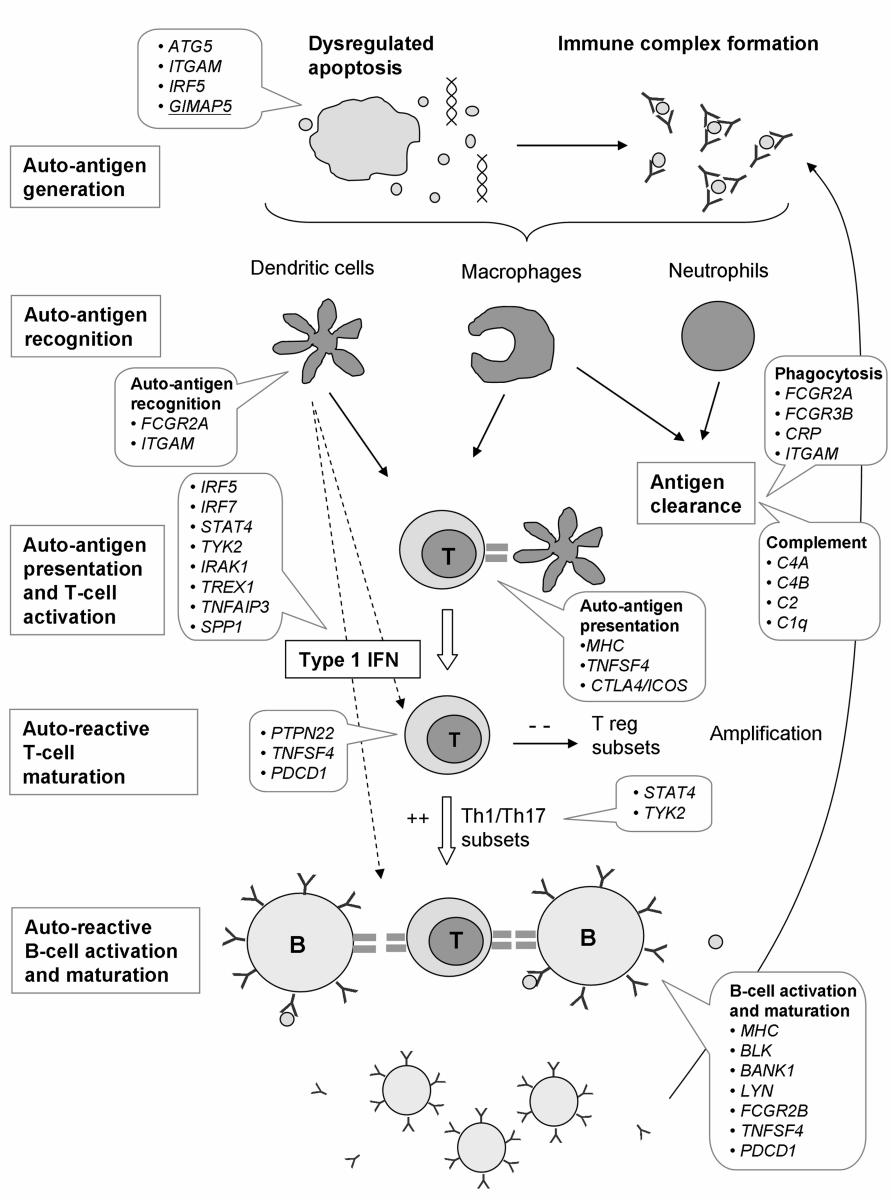

Figure 5. SLE susceptibility genes and their involvement in pathways important in the development of disease. Adapted from Rhodes and Vyse, The genetics of SLE: an update in the light of genome-wide association studies, Rheumatology, 2008, 47(11):1603-11, by permission from Oxford University Press.

2.3.3.1 Apoptosis, clearance and auto-antigen recognition

The importance of apoptosis and effective clearance of immune complexes is well recognized in SLE pathogenesis (reviewed in section 2.1.7.2) and many of the key SLE candidate genes play a role within these processes (Figure 5). Whether apoptosis itself

is dysregulated is less understood, however, several genes with suggested roles in apoptosis have been implicated in SLE susceptibility, including *ATG5*, *IRF5*, *ITGAM* and *GIMAP5*. Over-expression of *ATG5*, which is a regulator of autophagy and apoptosis, may result in accelerated apoptosis (Yousefi and Simon, 2007). *IRF5* promotes apoptosis upon signaling through tumour necrosis factor-related apoptosis-inducing ligand (TRAIL) death receptors (Hu and Barnes, 2009), while loss of *GIMAP5* function results in lymphocyte apoptosis (Pandarpurkar et al., 2003). Signaling through *ITGAM* has been shown to mediate neutrophil apoptosis (Coxon et al., 1996) and furthermore *ITGAM* has a role in immune complex clearance by phagocytosis (reviewed in (Harley et al., 2009)). Additional SLE susceptibility genes involved in clearance of immune complexes are *CRP* and genes encoding for the Fcγ-receptors. While CRP binding to FcγRIIA, which is expressed on macrophages and dendritic cells, leads to phagocytosis and release of inflammatory cytokines, CRP binding to FcγRIIB, which is expressed on neutrophils only, blocks the activating signals from this receptor (Li et al., 2009). It is not clear how the susceptibility variants in the genes involved in clearance contribute to disease. A plausible theory is that they affect the way that the encoded proteins bind immune complexes, which would result in impairing clearance (Harley et al., 2009; Rhodes and Vyse, 2008). Apoptotic cells that are not effectively removed are a source of autoantigens which will subsequently be recognized by APCs.

2.3.3.2 Type 1 interferons and immune response to infection

It has been shown also that patients with SLE have increased serum levels of the type 1 IFN cytokine IFN-α (Crow and Kirou, 2004) and furthermore that abnormally high serum IFN-α is a common heritable trait within SLE families in both healthy and affected members (Niewold et al., 2007). In addition, genes induced by the type 1 (IFN-α and -β) IFN pathway have been found to be up-regulated in individuals with SLE (Ronnblom and Pascual, 2008). IFN-α is important in maturation of dendritic cells and also in the production of pro-inflammatory cytokines and has diverse effects on immune functions, including stimulation of Th1 pathways, promotion of B-cell activation and regulation of apoptosis (Ronnblom and Pascual, 2008). In SLE, the type I IFNs can contribute to loss of tolerance and activation of autoreactive T- and B-cells with production of autoantibodies (Kaplan, 2005). Studies on genes in this pathway have identified *IRF5*, *TYK2* and *STAT4* as risk genes in SLE susceptibility ((Remmers et al., 2007; Sigurdsson et al., 2005), Figure 5). The association identified to *KIAA1542* (Harley et al., 2008) might be an indirect association to the adjacent *IRF7* gene, which is thought to have similar functions as *IRF5*. Furthermore, the SLE susceptibility genes *SPP1 (Osteopontin)* and *TREX 1* are also suggested to be involved in the type I IFN pathway (reviewed in (Moser et al., 2009), Figure 5). SPP1 is a multifunctional cytokine that is over-expressed in SLE patients and shown to be critical for IFN-α production in murine plasmacytoid dendritic cells (pDC) (Kariuki et al., 2009). *TREX1* encodes for a DNA exonuclease, which has a role in digesting cytosolic DNA and thereby prevents activation of a cell-intrinsic type I IFN response pathway (Stetson et al., 2008).

The majority of cells produce Type 1 IFNs as part of their early response to viral infection. In SLE patients production of IFNs, especially in pDCs, is also induced by endogenous derived nucleic acids (Ronnblom and Pascual, 2008). Type 1 IFNs have multiple effects in the immune system and exert their functions through the IFN receptor (IFNAR) on target cells, which induce production of proteins that inhibit viral replication. The IFNAR consists of two subunits, IFNAR1 and IFNAR2, and binding of type 1 IFN results in activation of the Janus kinases JAK1 and TYK2 with a subsequent recruitment and activation of several signal transducers and activators of transcription (STAT1-6, IRF5 and IRF7), finally leading to transcription of IFN signature genes (Bengtsson et al., 2000; Decker et al., 2002; Hardy et al., 2004; Kozyrev and Alarcon-Riquelme, 2007; Richter et al., 1998).

Furthermore, the toll-like receptors (TLRs) mediate type I IFN gene transcription (David, 2002). TLRs are pattern recognition receptors which are expressed on APCs that bind microbial components and thereby provide activating co-stimulatory signals. In the context of SLE, TLR7 and TLR9 are of particular interest because of their ability to recognize nucleic acids. Activation of these two, as well as TLR4, leads to phosphorylation of several transcription factors, including *IRF5* (Honda and Taniguchi, 2006). The newly identified SLE susceptibility gene *TNFAIP3* encodes for an enzyme involved in the downstream signal transduction of TLRs and is negative regulator of the nuclear factor-κB (NF-κB) pro-inflammatory responses following signal transduction and thus it plays a key role in modulating a broad range of cellular functions, including cell activation and cytokine signaling (Graham et al., 2008). Similar to *TNFAIP3*, the *IRAK1* gene encodes for a serine/threonine protein kinase involved in TLR signaling. Recent studies in murine models of SLE support a possible role for *IRAK1* in induction of IFN-α/γ, regulation of NF-κB in T-cells and TLR activation (Jacob et al., 2007; Jacob et al., 2009).

2.3.3.3 Auto-antigen presentation, T-cell activation and maturation

As reviewed in section 2.1.7.1 pathogenic autoantibodies are produced by B-cells in the presence of stimulating antigen, which does require co-stimulation by CD4+ T-cells. The importance of this pathway in SLE susceptibility is implicated by association to several genes, including the *MHC* (section 2.3.1), *TNFSF4*, *CTLA4-ICOS*, *PTPN22, STAT4* and *TYK2*, suggested to be involved in either APC/T-cell cross-talk, T-cell activation and/or maturation (reviewed in (Criswell, 2008; Harley et al., 2009; Moser et al., 2009; Rhodes and Vyse, 2008), Table 2, Figure 5). *PDCD1* (section 2.3.1) also falls under this category, given its inhibitory function on T-cell proliferation and cytokine secretion. Studies have implicated that *TYK2* and *STAT4* (section 2.3.3.2) in addition to type 1 IFN, also respond to IL-12 and IL-23 stimulation (reviewed in (Kaplan, 2005; Schindler and Plumlee, 2008)), which leads to activation of *STAT4* and subsequently to production of IFN-γ and IL-17. These cytokines are important in directing Th1 as well as a Th17 responses (reviewed in (Korman et al., 2008)

TNFSF4 (also known as OX40L) is expressed on antigen-presenting cells, while its receptor, TNFSFR4, is expressed on activated CD4+ T-cells (reviewed in (Moser et al., 2009)). The binding of TNFSF4 to the TNFSF4 receptor on T-cells provides a co-stimulatory signal which mediates proliferation and expansion of T-cells during the primary immune response. A haplotype in *TNFSF4* has been found associated with SLE risk as well as up-regulated *TNFSF4* transcription and increased lymphocyte surface expression of TNFSF4 receptor (Cunninghame Graham et al., 2008). The association with SLE within the *CTLA4-ICOS* lies in the 3'untranslated region (UTR) of *CTLA4* and recent studies have suggested that this association may extend into the promoter of the neighboring gene, *ICOS* (Rhodes and Vyse, 2008). *CTLA4* is expressed on activated T-cells and has the ability to down-regulate these after activation, while *ICOS* is involved in the maintenance of T-cell activation and is up-regulated in patients with SLE.

The *PTPN22* gene encodes for the lymphoid-specific tyrosine phosphatase (LYP) protein, which modulates signal transduction in T-cells. A functional SNP (R620W) within *PTPN22* has been associated to SLE (Harley et al., 2008; Kyogoku et al., 2004). The risk allele (620W) is thought to increase catalytic activity, compared with the non-risk variant (620R), and to be a more potent suppressor of T-cell receptor signaling (reviewed in (Rhodes and Vyse, 2008)).

2.3.3.4 B-cell activation and maturation

Auto-reactive B-cells are the source of the high-affinity autoantibodies responsible for SLE pathogenesis. Several genes involved in the B-cell pathway and particularly B-cell receptor (BCR) signaling, including the *MHC* (section 2.3.1), *BLK*, *LYN*, *BANK1* and *FCGR2B*, have been shown to contribute to SLE susceptibility and thereby illustrates the importance of this pathway in disease pathogenesis (reviewed in (Criswell, 2008; Harley et al., 2009; Moser et al., 2009; Rhodes and Vyse, 2008) Table 2, Figure 5).

BANK1 encodes for a B-cell adapter protein expressed predominantly in B-cells and is thought to alter B-cell activation thereby increasing the risk of SLE (Kozyrev et al., 2008). LYN is protein tyrosine kinase that physically associates with the BCR, together with its binding partner BANK1, while BLK affects functions associated with the development of B-cells before the appearance of the BCR. Preliminary data suggest that SLE risk is associated with reduced BLK transcription (Hom et al., 2008). LYN, BANK1 and BLK all play a critical role in controlling the activation of B-cells following signaling through the BCR (reviewed in (Rhodes and Vyse, 2008)). After ligand binding and BCR aggregation, an early intracellular event is the recruitment and activation of Src-family protein tyrosine kinases, including BLK and LYN, which mediate further intracellular signaling. BANK1 is a scaffold protein that acts downstream of LYN and facilitates release of intracellular calcium that is an important event in B-cell activation.

Furthermore, the previously described genes *PDCD1* and *TNFSF4* (section 2.3.1 and 2.3.3.3), probably play a complex role in both B- and T-cell activation through a bidirectional signaling mechanism (reviewed in (Moser et al., 2009; Rhodes and Vyse, 2008)). B-cells express *TNFSF4* and it has been shown that interaction of TNFSF4 with its receptor can induce B-cell activation and differentiation.

2.3.3.5 Genes of unknown function

Several genes with unknown function and with no apparent relationship to the immune system, including *ICA1*, *SCUBE1*, *MNAT2*, *KIAA1542*, *PXK*, *XKR6*, *FAM167A*, was identified in the recent GWA studies (reviewed in (Moser et al., 2009)). The mechanisms by which these genes contribute to risk of SLE are yet to be discovered. Genes with unknown function have the potential to reveal new pathways involved in SLE susceptibility or interaction with already known pathways that would not have been identified otherwise. *MAMDC1*, identified though positional mapping (Paper II), also belong to this category of genes with unknown function.

2.3.4 Shared susceptibility with other autoimmune diseases

In addition to the *MHC*, many of the SLE susceptibility genes have been found associate to other autoimmune diseases (reviewed in (Gregersen and Olsson, 2009; Zhernakova et al., 2009)). For example, *PTPN22* is associated with rheumatoid arthritis (RA), Graves' disease, type 1 diabetes (T1D) and, autoimmune thyroid disease; *TNFAIP3* with Cohn's disease; *STAT4* with RA; and finally *CTLA4* is associated with T1D and RA. Thus it is obvious that some genetic variants clearly predispose to multiple autoimmune diseases, suggesting that many of these diseases share common pathways of pathogenesis, despite their highly heterogeneous clinical manifestations. However, the lack of such overlap for some genes provides evidence that distinct mechanisms also exist and where the genes involved in B-cell dysregulation appear to be specific for SLE (reviewed in (Rhodes and Vyse, 2008)).

3 AIMS OF THE THESIS

The overall aim of this thesis was to study susceptibility genes in SLE using three different approaches:

I Evaluate the role of *GIMAP5* in SLE susceptibility, which was achieved by a candidate gene association study and subsequent gene characterization (Paper I).

II Positional mapping and gene identification of susceptibility gene(s), with subsequent gene characterization, within our previously identified susceptibility loci on chromosomes 14q21-q23 (Paper II).

III Study whether previously identified SLE susceptibility genes constitute a risk in Finnish patients and if a combination of associated genes modifies epistatically the risk of SLE (Papers III, IV and V).

4 MATERIAL AND METHODS

4.1 SAMPLES

4.1.1 Finnish family cohort (Papers I–V)

The Finnish family cohort comprises 192 families, with a total of 236 SLE patients and their healthy relatives, and has been the basis of all studies in this thesis. The recruitment of this cohort started in Finland in 1995 and approximately 80-85% (1200 out of 1500 available patients based on the reported prevalence in Finland (Helve, 1985)) of all Finnish SLE patients requiring hospital-based treatment were contacted. All patients were interviewed by the same physician, and the case records from the hospitals, where the patients were treated, were reviewed. Patients with a positive family history of SLE and fulfilling the ACR criteria for the classification of SLE (Tan et al., 1982), were asked to participate in the study together with their unaffected and/or affected family members. Blood samples were obtained from a total of 252 families, of which 53 were multiply affected by SLE and the remaining were families of sporadic patients. The clinical characteristics of the 236 patients included in Papers I–V are described in Table 3. For a more detailed description of the collection, subphenotyping and clinical characteristics of this cohort see (Koskenmies et al., 2004b; Koskenmies et al., 2001).

4.1.2 Finnish case-control cohort (Paper II, III and V)

The Finnish case-control cohort, which consists of 86 SLE cases and 356 controls, was used for replication in Paper II and in combination with the probands from the Finnish family cohort in Papers III and V (further described in section 4.3.3). All patients with clinical diagnosis of SLE and attending the Departments of Dermatology at Helsinki and Tampere University Central Hospitals during 1995-2005 were identified from the corresponding hospital registries, and contacted by mail or phone. The presence of correct clinical diagnosis, defined as the fulfillment of the ACR criteria for the classification of SLE (Tan et al., 1982), was confirmed from the patients' hospital records and diagnosis of SLE was further verified by a rheumatologist. Unaffected unrelated family members (spouses or common-law spouses) were asked to participate in the study as control individuals, and an existing collection of unrelated healthy individuals was also used as control samples. The clinical characteristics for all patients included in this cohort are described in Table 3, while a more detailed description of this cohort is found in (Koskenmies et al., 2008).

4.1.3 British family cohort (Papers I and II)

The British family cohort is a large collection of SLE nuclear families from the UK and has been used for replication of our initial findings in Papers I and II. The cohort predominantly consists of one affected offspring per family and the collection of this material is described in detailed in (Russell et al., 2004). The diagnosis of SLE was

established by a telephone interview, health questionnaire and details from clinical notes and all patients confirmed to the ACR criteria for SLE (Tan et al., 1982). In total, the British cohort comprises of 549 patients of European-Caucasian (EC) descent, 37 patients of Indian-Asian (IA) descent, 31 patients of Afro-Caribbean (AF) descent and 12 patients with mixed ethnicity, and their unaffected family members. The clinical characteristics for all patients included in this cohort are described in Table 3.

4.1.4 Swedish case-control cohort (Paper II)

The Swedish case-control cohort consists of 304 cases and 307 controls and was used for replication in Paper II. All patients in this cohort were interviewed and examined by a rheumatologist at the Department of Rheumatology, Karolinska University Hospital (Svenungsson et al., 2003) and all fulfilled the ACR criteria (Tan et al., 1982). The control samples were collected from population-based control individuals and individually matched for age and gender. The clinical characteristics for all patients included in this cohort are described in Table 3.

Table 3. Clinical characteristics for all patients included in this thesis.

	Finnish family cohort	Finnish case-control cohort	British family cohort				Swedish case-control cohort
Ethnicity	EC	EC	EC	IA	AF	MIX	EC
Number of patients	236	86	549	37	32	12	304
Females	94	93	92	86	100	100	90
Mean age at onset (range)	29 (1-66)	31 (8-73)	26 (3-53)	29 (27-45)	25 (9-35)	21 (16-27)	n.a
Mean age at diagnosis (range)	33 (6-72)	35 (13-76)	30 (10-54)	30 (27-45)	36 (28-47)	21 (16-28)	31 (7-74)
Butterfly rash	51	74	85	77	78	50	52
Discoid rash	10	41					17
Photosensitivity	69	80	75	57	44	20	52
Mouth ulcers	18	16	73	63	71	0	34
Arthritis	83	64	72	80	69	50	87
Pleuritis	18	n.a.	30	37	16	33	40
Pericarditis	16	n.a.					19
Nephritis	30	20	33	51	44	50	41
Convulsions	5	n.a.	18*	19*	10*	0*	n.a.
Psychosis	1	n.a.					n.a.
Leukopenia	68	37	n.a.	n.a.	n.a.	33	50
Thrombocytopenia	16	16	23	11	13	33	21

All values are presented as % over available values, n.a. not available, * Also encompasses serious depression, EC, European-Caucasian, IA, Indian-Asian, AF, Afro-Caribbean and MIX, mixed ethnicity.

4.1.5 Ethical aspects

All participants included in this thesis gave written informed consent for participation in genetic studies on SLE and the study protocols were reviewed and approved by the ethical committee at Karolinska Institutet; The Ethical Review Boards of Helsinki and Tampere University Central Hospitals; the ethical committee at University of Helsinki; and the Multi-Centre Research Ethics Committee. All studies were conducted according to the Declaration of Helsinki ethical principles for medical research involving human subjects.

4.2 GENOTYPING (PAPERS I–V)

4.2.1 Microsatellites (Papers II and V)

Microsatellite genotyping, using the MegaBACE™1000 Genotyping System (GE Healthcare), was performed for the fine mapping of the chromosome 14q21-q23 region (Paper II) and for the genotyping of the *IRF5* CGGGG indel (Paper V). Genomic DNA was PCR-amplified using fluorescently labeled primers, multiplexed (Paper II), separated using capillary array electrophoresis and analyzed using the MegaBACE™ Genetic Profiler v2.0 software (see (Koskenmies et al., 2004a) for a detailed protocol).

4.2.2 SNPs (Papers I–V)

All genotyping methods used in this thesis were based on enzyme-assisted single nucleotide primer extension (reviewed in (2006; Syvanen, 2001), Figure 6). In Papers I-III and V the Sequenom (Sequenom Inc.) matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry method was used, with the exception of the genotyping of the Finnish family cohort in Paper I where the MegaBACE™ 1000 single-nucleotide primer extension (SNuPE) method (GE Healthcare) was used (see paper for protocol). In Paper IV the SNPstream genotyping system (Beckman-Coulter Inc) was used (see paper for protocol and (Bell et al., 2002)).

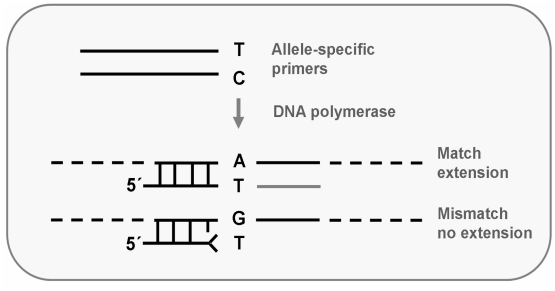

Figure 6. Allele-specific primer extension in which two primers complementary to the allelic variant at their 3′-end are anneal to their target sequence adjacent to the SNP and where primers with perfectly matched 3′-ends will be extended. Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Genetics (Syvanen, 2001), copyright 2001.

The Sequenom MALDI-TOF mass spectrometry method uses either MassEXTEND® (hME) (Paper I and II) or iPLEX (Paper II, III and V) methods for allele-specific primer extension (Jurinke et al., 2002). The protocol for the hME genotyping is described in the supplementary material of Paper I. The iPLEX genotyping amplification reactions were run with similar PCR protocols, but with 10 ng of

genomic DNA, 100 nM of primer mix, 500 mM of dNTP mix, 1.625 mM of MgCl2 and 0.5U of HotStarTaq DNA Polymerase. All PCR and extension assays were designed using SpectroDESIGNER software (Sequenom Inc.). Unincorporated dNTPs were dephosporylated by addition of 0.3 U shrimp alkaline phosphatase enzyme to each sample. Extension reactions were then conducted in a total volume of 9 µl using 0.625 µM for low mass primers and 1.25 µM for high mass primers and the Mass EXTEND Reagents Kit before being cleaned using SpectroCLEANER (Sequenom Inc.). Desalted primer extension products were analyzed by a MassARRAY mass spectrometer (Bruker Daltonik). The resulting mass spectra were analyzed for peak identification using the SpectroTYPER RT 3.3.0 software for iPLEX assays (Sequenom Inc.). All genotypes were independently verified by two investigators.

To assure for genotyping consistency and quality each genotyped assay was initially validated by comparing genotype concordance from the genotyping on our platform in a set of 14 CEU trios (CEPH Utah residents with ancestry from northern and western Europe), with genotypes available through the HapMap consortium (www.hapmap.org). Furthermore, internal concordance was analyzed on 14 additional unrelated individuals of Caucasian descent. Hardy-Weinberg equilibrium (HWE) was analyzed in a total of 55 unrelated individuals to ensure that each marker was in equilibrium. Also the percentage of negative controls with genotypes for each assay was accounted for, where assays yielding higher than 50% were excluded. Success rate of all assays was required to be at least 85%. Assays not fulfilling the quality criteria were excluded from further genotyping. In the following genotyping, 90 samples from each individual sample set were genotyped twice and analyzed for concordance to assure for genotyping consistency. In addition to analyzing HWE, negative controls with genotypes and success rate as in the validation step, PedCheck was used to detect Mendelian inconsistencies when applicable (O'Connell and Weeks, 1998) and markers were excluded if they yielded more than 10% errors.

4.3 ASSOCIATION ANALYSIS (PAPERS I–V)

4.3.1 Haplotype pattern mining (Paper II)

Haplotype pattern mining (HPM) is a data-mining-based association analysis method, based on the discovery of recurrent marker patterns (Sevon et al., 2001; Toivonen et al., 2000). The HPM algorithm searches for haplotype patterns in case and control chromosomes and sorts these by their strength of association to the trait (i.e. more or less frequent in cases). A non-parametric model is then used for localizing the underlying locus. HPM has been shown to be robust and powerful for sparse marker maps and allows for missing or erroneous data and was used for the initial fine mapping of the linkage region on chromosomes 14q21-q23 in Paper II. The data was analyzed using HPM v. 2.0 and obtained *P*-values were permutated 50,000 times to compensate for variable marker densities and marker information content.

4.3.2 Transmission and Pedigree disequilibrium test (Papers I, II and IV)

TDT (Spielman et al., 1993), as implemented in the software packages GENEHUNTER 2.1 (Kruglyak et al., 1996) or UNPHASED (Dudbridge, 2003, 2008), was used for the analysis of single markers and haplotypes in Papers I and IV, respectively. The TDT compares the frequencies of transmitted vs. untransmitted alleles in the affected offspring, by using the untransmitted parental alleles as controls (see section 2.3.3). Association between *STAT4* variation and specific SLE phenotypes in parent-affected offspring trios for a specific phenotype were further analyzed in Paper IV. Single marker association in Paper II was analyzed using PDT (Martin et al., 2000), which is an extension of the TDT that integrates extended family information (see section 2.3.3), and PDTPHASE (Dudbridge, 2003), which is an extension of the PDT. In Paper II, haplotypes were analyzed using the "haplo.stats 1.3.0" software from R (www.r-project.org). LD among the genotyped SNPs in Papers I, II and IV was visualized using different versions of the Haploview software (www.broad.mit.edu/mpg/haploview) (Barrett et al., 2005). For association analysis using discordant sib pairs in Paper I, we used the discordant allele test (Boehnke and Langefeld, 1998). In Paper I, global significance of the distribution of haplotypes was assessed by a randomisation test on the transmitted alleles/haplotypes using 50,000 iterations. In Paper IV, which was a replication study, nominal *P*-values <0.05 were considered significant. The correction for multiple testing in Paper II is described in section 4.3.4.

4.3.3 Case–control analysis (Papers II, III and V)

In Paper II, COCAPHASE, which is a part of the UNPHASED software (Dudbridge, 2003), were used for the single marker analysis in the case-control cohorts (see section 4.3.4 for correction for multiple testing) and haplotypes were analyzed using the "haplo.stats 1.3.0" software from R (www.r-project.org). In Papers III and V, one proband from each SLE family (section 4.1.1) was included in the analysis together with the Finnish sporadic SLE cases and controls (section 4.1.2). Single SNP and haplotype associations were investigated using a chi-square test as implemented in the Haploview program v. 4.0 (Barrett et al., 2005). As these were a replication studies, nominal P-values <0.05 were considered significant. In Paper III, the association between risk-allele carrier status and SLE phenotypes was examined using a chi-square or Fisher exact test, when appropriate. Analyses were done with SPSS v. 15.0 and *P*-values <0.05 after Bonferroni multiple testing correction were considered significant.

4.3.4 Meta analysis (Papers I and II)

In Paper I, a meta-analysis (Lohmueller et al., 2003) of the independent sample sets was performed in two steps, first using the Breslow–Day test for non-compatibility of ORs and then the Mantel–Haenzel method as implemented in the R software (www.r-project.org) for pooled estimate of ORs. In Paper II, meta-analysis of the case-control and family data was performed using the Kazeem and Farrell (Nicodemus, 2008) fixed effect model implemented in the R package catmap1.5. To take into account multiple testing in Paper II, the nominal significance threshold of *P*-value = 0.05 was corrected

by finding the number of independent SNPs, using a Principal Component Analysis of the SNPs correlation matrix (Nyholt, 2004).

4.3.5 Interaction and additive joint effect analysis (Paper II, III and V)

Multiple logistic regression models were used to estimate the interactive effects of SNPs by adding an interaction term between the genotypes of interest, using either the Stata software v. 8.0 (Paper III) or the R software (Papers II and V). In Paper III, a logistic regression model was also used to estimate the additive joint effects, using SPSS v. 15.0.

4.4 SEQUENCING (PAPERS I AND II)

For the identification of novel and functional SNPs in *GIMAP5* and *MAMDC1*, we sequenced non-repetitive genomic DNA spanning the entire genomic regions of the two genes. Primers for all assays were designed using Primer3 (frodo.wi.mit.edu/primer3/). The amplification reactions were run using a standard PCR protocol and followed by sequencing reactions performed on both the forward and reverse strand. Purified sequencing products were separated using the fluorescence-based MegaBACE™1000 Automated Capillary DNA Sequencing System (GE Healthcare) and visualized using the MegaBACE™ Sequencer Analyser software 3.0. The sequences were then analyzed by comparing with public sequences obtained from either NCBI (www.ncbi.nlm.nih.gov) or UCSC (www.genome.ucsc.edu), using the Staden Package computer Programs Pregap4 and Gap4 (www.gap-system.org) or the BLAST databases (www.blast.ncbi.nlm.nih.gov/Blast.cgi). A detailed sequencing protocol can be found in the supplementary material of Paper I.

4.5 GENE EXPRESSION ANALYSES (PAPERS I AND II)

4.5.1 Northern blot (Papers I and II)

Northern blotting (Alwine et al., 1977) is a useful method for studying alternative RNA transcription as well as expression and was used for this purpose in Papers I and II. Complimentary DNA (cDNA) probes specific to *GIMAP5* or *MAMDC1* were labeled with P^{32}-dCTP (GE Healthcare) by random priming and hybridized to commercial human multiple tissue polyA+ RNA Northern blots according to the manufacturer's (Ambion, Clontech and OriGene Technologies), instructions and β-Actin cDNA was used for normalization. Although this method can be used for quantification of RNA by comparing RNA levels between multiple samples on a single membrane, Northern blotting lacks the accuracy of quantitative RT-PCR.

4.5.2 Real-time and quantitative real-time PCR (Papers I and II)

Real-time PCR (RT-PCR) and quantitative real-time PCR (qRT-PCR) are methods used to study gene expression and consist of three phases: the exponential phase, the linear phase and the plateau phase. Initially, when a PCR reaction is not limited by enzymatic activity or substrates, product generation is exponential and has close to

100% efficiency. As the reagents eventually become depleted the reaction will reach a plateau.

In an end point RT-PCR approach, cDNA is amplified for a certain number of cycles and visualized through agarose gel electrophoresis. Given that this method measures the amount of cDNA after a fixed number of cycles where, theoretically, all of the samples could have reached the same total amount of amplified DNA, this is not an optional method for quantification of cDNA. In qRT-PCR, the quantification of PCR products is studied in "real time" during each PCR cycle, yielding a quantitative measurement of PCR products accumulated during the course of the reaction. The quantification could be measured either by absolute or relative levels. The most common method for relative quantification is the $2^{-\Delta\Delta C}{}_T$ method, where the C_T is the number of cycles needed to reach an arbitrary horizontal threshold, where all samples are exponentially quantified (Livak and Schmittgen, 2001) and relies on the assumption of 100% efficiency and the presence of an endogenous control that is expressed at a constant level between samples (VanGuilder et al., 2008). In this method ΔC_T is calculated as the difference in C_T values for the gene of interest and the endogenous control for each sample. The $\Delta\Delta C_T$ is then calculated by subtracting the control ΔC_T from the ΔC_T calculated for treated samples. The negative value of this subtraction, the $-\Delta\Delta C_T$, is used as the exponent of 2 in the equation and represents the relative difference compared to the control sample. The exponent conversion is based on the assumption of 100% efficiency, i.e. that the reaction doubles the amount of product per cycle.

The end point RT-PCR approach was used for the identification of tissues and cell-lines expressing either *GIMAP5* or *MAMDC1* (Papers I and II); to investigate expression of alternative *MAMDC1* transcripts (Paper II); and to investigate differential termination of *GIMAP5* mRNA transcription (Paper I). Quantitative RT-PCR was used to investigate several different aspects of gene expression (Papers I and II), including identification of suitable cell-lines with sufficient mRNA expression for qRT-PCR experiments; to investigate expression of alternative transcripts (Paper II); to investigate the effects exerted by a number of cytokines with important roles in inflammation on mRNA expression in monocytes; and comparison of mRNA expression between patients (n = 9) and controls (n = 9). Two different methods for qRT-PCR were used: the TaqMan qRT-PCR (Applied Biosystems) method (Paper I), which uses an amplicon-specific fluorescent labelled probe to measure amplified PCR product (VanGuilder et al., 2008), and the SYBR green qRT-PCR (Applied Biosystems) method (Paper II), which instead uses intercalating dyes (VanGuilder et al., 2008). Both methods use cDNA as a template and all reactions were performed in triplicates with the 7500 Fast Real-Time PCR system using standard protocols for either TaqMan or SYBR green (Applied Biosystems) and *GAPDH* as the endogenous control. Relative expression was compared either to a randomly chosen reference sample or unstimulated cells and was calculated using the $2^{-\Delta\Delta C}{}_T$ method. Detailed protocols are given in each respective paper.

4.5.3 Allelic expression (Paper I)

Allele-specific mRNA expression levels of *GIMAP5*, relative to rs759011, rs1046355, rs10361, rs6598 and rs2286899, were assessed in patients (n = 9) and controls (n = 9) by sequencing (see section 4.4) and subsequent comparison of peak heights between individual cDNA to genomic DNA, with subsequent calculations of allele ratio (described in (Pastinen et al., 2004)). The cDNA ratio values were normalized by dividing with the genomic values and the data were pooled by genotype (risk heterozygotes vs. non-risk heterozygotes) to evaluate whether the normalized value differed from equal expression.

4.6 PROTEIN EXPRESSION

4.6.1 Western blot (Paper I)

To test the specificity of the GIMAP5 polyclonal antibody, a Western blot analysis was performed. In this method proteins are separated on a denaturing SPS-PAGE gel and transferred to a membrane, where they target protein is subsequently detected using an primary antibody and an secondary antibody conjugated with an enzyme such as alkaline phosphatase or HRP .

4.6.2 Immunohistochemistry (Papers I and II)

Immunohistochemistry (IHC) analyses were performed to study the expression of the GIMAP5 and MAMDC1 proteins in multiple human tissue sections. An affinity purified rabbit polyclonal antibody generated against the GIMAP5 exon-3-specific peptide LGREREGSFHSNDLF (Sigma Genosys) or a commercial MAMDC1 rabbit polyclonal antibody (Atlas Antibodies) was used for protein detection. There are numerous methods used for IHC in which fluorescent dye, enzyme, radioactive element or colloidal gold can be used for visualization. In Paper I we used the three-step avidin–biotin–complex (ABC) technique for detection of GIMAP5, in which a biotinylated secondary antibody interacts with complex of avidin-biotin peroxidase (see Paper I for protocol). For detection of MAMDC1 in Paper II we used the two-step polymeric technique, in which a horseradish peroxidase (HRP) labelled polymer is conjugated directly with the secondary antibody. For visualization, diaminobenzidine was used as a chromogenic substrate, which produces a brown end product (see Paper II for protocol)

5 RESULTS AND DISCUSSION

5.1 PAPER I – A FUNCTIONAL CANDIDATE GENE STUDY OF *GIMAP5* IDENTIFIED AN ASSOCIATION BETWEEN A COMMON HAPLOTYPE AND INCREASED RISK OF SLE

In Paper I we used a functional candidate gene approach to investigate the role of *GIMAP5* in SLE susceptibility. The rationale behind this was that *GIMAP5* had been shown to protect cells from apoptosis, following the finding that a protein truncating frame-shift mutation in the rat homolog of this gene (Gimap5, previously Ian5 and Ian4) was the cause of lymphopenia in the diabetes-prone Biobreeding (BB) rat (Hornum et al., 2002; MacMurray et al., 2002). Several studies have since then shown that intact GIMAP5 is necessary for the survival of peripheral T-cells (Pandarpurkar et al., 2003; Pino et al., 2009; Schulteis et al., 2008), yet the exact mechanism by which loss of GIMAP5 induces apoptosis is not clear. Seeing that lymphopenia is one of the haematological manifestations in SLE (Table 1), that a possible dysregulation of apoptosis is contributing to the pathogenesis of SLE (2.1.7.2) and furthermore that *GIMAP5* is localized within a suggestive linkage region on 7q36 (Gaffney et al., 2000); *GIMAP5* was considered a potential candidate gene also in SLE.

By analyzing seven SNPs in the Finnish and British family cohorts, capturing all common variation within *GIMAP5*, we found a significant increase in SLE risk associated with the most common haplotype of this gene (*P*-value = 0.003, OR = 1.260; Table 2 and Figure 2 in Paper I) and that the risk was increased in families with probands diagnosed with trombocytopenia (*P*-value = 0.015, OR = 2.110; Figure 2 in Paper I). The strongest association was seen in the British Asian trios, which in retrospect is very interesting since a common *GIMAP5* haplotype was recently found to be associated with susceptibility to SLE in a Korean case-control study (Lim et al., 2009). Two of the SNPs included in our study were identical to that of the Korean study and importantly the same alleles of these two SNPs (rs1046355 C and rs6598 G) were included in the common risk haplotype in both studies. The non-replication of *GIMAP5* in the recent GWA studies (Graham et al., 2008; Harley et al., 2008; Hom et al., 2008; Kozyrev et al., 2008), which all were performed in individuals of European descent, may suggest that *GIMAP5* is a more potent risk factor in individuals of Asian descent.

Given that *GIMAP5* is essential for lymphocyte survival, we wanted to know, whether the mRNA expression of this gene differs between cases and controls. When comparing mRNA expression in peripheral blood mononuclear cells from nine patients with SLE and nine healthy control individuals we saw a trend for lower expression in patients (Figure 4C in Paper I). Interestingly, *GIMAP5* was recently found down-regulated in T regulatory (T reg) cells from T1D subjects (Jailwala et al., 2009), thus differential expression may be one important aspect for the contribution of *GIMAP5* to autoimmunity.

We could not identify any coding non-synonymous variation within *GIMAP5*. However, the G allele of rs6598 was found to introduce a switch from a canonical AATAAA polyadenylation (polyA) signal to a rare (AATAGA) polyA signal and thus hypothesised to cause a differential termination of the transcription of *GIMAP5* (Figure 1A in Paper I). When comparing differences in mRNA lengths between individuals homozygous for the rarer AATAGA signal with individuals heterozygous or homozygous for the conservative polyA signal AATAAA, we found that individuals homozygous for the rarer AATAGA signal had a higher amount of longer mRNA transcripts (Figure 1C in Paper I). Interestingly, rs6598 was later shown to be associated with the presence of high levels of IA-2 autoantibodies in newly diagnosed T1D patients (Shin et al., 2007). Differential transcript length may affect stability or translatability of mRNA transcripts and thereby affects levels of the translated protein (Edwalds-Gilbert et al., 1997).

5.2 PAPER II - POSITIONAL MAPPING OF THE CHROMOSOME 14Q21-Q23 LINKAGE REGION IDENTIFIED *MAMDC1* AS A CANDIDATE GENE IN SLE

The region on chromosomes14q21-q23 had showed suggestive linkage in a previous genome-wide scan including a subset of the Finnish family cohort (NPL score = 2.200, *P*-value = 0.020) (Koskenmies et al., 2004a). Additional genotyping identified sharing of two common haplotypes in this region, spanning microsatellite markers D14S978-D14S589-D14S562 and D14S1009-D14S748 (*P*-values = 0.006 and 0.140, respectively) (Koskenmies et al., 2004c). Two other independent studies had also identified suggestive linkage to this region (Gaffney et al., 1998; Shai et al., 1999) and thus this region was considered to be of particular interest for gene identification in Paper II.

Fine mapping was first performed in our Finnish family cohort, focusing on the two regions that previously showed haplotype sharing (Koskenmies et al., 2004a; Koskenmies et al., 2004c). To increase our chances of gene identification we used both HPM and PDT (see section 4.3) in the analysis and could identify one region, spanning markers D14S1068 and rs1955810, showing positive signals of association with both method (*P*-value $\leq$0.050, Figure 1 in Paper II). To our excitement, this region contained only two genes: *MAMDC1* and *RPL10L* (ribosomal protein L10-like protein), with the

distances to nearby genes being larger than 1 Mb in both directions (Figure 1 in Paper II). Since most of the associated markers mapped within *MAMDC1*, we focused our downstream analysis to this gene and for this purpose the Finnish and Swedish case-control cohorts and a subset of the British family cohort were recruited to the study (see section 4.1). Subsequently, we genotyped 24 SNPs in the four sample populations and significant associations were identified for the SNPs rs961616 (*P*-value = 0.001, OR = 1.292) and rs2297926 (*P*-value = 0.003, OR = 1.349) in a combined analysis (Figures 1 and 2 and Supplementary Table 2 in Paper II).

The association to *MAMDC1* gene (also known as *MDGA2*) is intriguing since this gene does not appear to function within any of the pathways thought to be of importance in SLE susceptibility and may thus reveal new insights in SLE pathogenesis. The *MAMDC1* gene spans 835 kb of DNA sequence on the reverse strand of chromosome 14q21.3 and is predicted to be composed of two alternative first exons (1a and 1b) and 16 downstream exons giving rise to two mRNAs of similar size (5375 bp for isoform 1, and 5239 bp for isoform 2, respectively; Figure 3A in Paper II). By Northern blotting, using a probe specific for isoform 1, we could show expression of a mRNA transcript corresponding to 5 kb in several tissues, as well as several additional splice variants of varying sizes ranging from less than 1 kb to over 9 kb (Figure 5 in Paper II). Furthermore, by designing primers specific to isoforms 1 and 2 and performing RT-PCR and qRT-PCR expression studies, we found that isoform 2 was only expressed in testis, while isoform 1 was found expressed in several tissues and cell-lines (data not shown). Isoform 1 corresponds to the MAMDC1 full-length protein, composed of 956 amino acids and containing 6 immunoglobulin (Ig) like domains, a fibronectin type III like (FNIII) fold domain, and a meprin/A5-protein/PTPmu (MAM) domain, as well as a C-terminal GPI anchoring signal and an N-terminal signal sequence (Figure 3B in Paper II). This domain structure is identical to that of the rat orthologs MDGA1 and MDGA2, both proposed as members of a novel subgroup of the Ig cell adhesion molecule (CAM) super family involved in cell-cell adhesion, migration and the development of neuronal connections (Litwack et al., 2004; Walsh and Doherty, 1997).

Remarkably, *MAMDC1* was recently found associated with neuroticism (van den Oord et al., 2008), which is a trait that reflects a tendency toward negative mood states and is linked to internalizing psychiatric conditions, such as anxiety and depression (Brandes and Bienvenu, 2006; Widiger and Trull, 1992). Furthermore, *MAMDC1* exonic CNVs were recently shown to contribute to risk in autism spectrum disorders (Bucan et al., 2009). In addition, the *MAMDC1* homolog *MDGA1* has been found to associate with schizophrenia (Kahler et al., 2008). Thus, a role for *MAMDC1* in the CNS seems probable. With regard to SLE, this finding is interesting, since neuropsychiatric manifestations are among the ACR criteria ((Tan et al., 1982), Table 1), but unfortunately we were underpowered to test for any association between *MAMDC1* variation and neuropsychiatric manifestations. It may be worth mentioning that one of our associated SNPs, rs2297926, is predicted to affect the binding of the transcription factor Engrailed. This transcription factor has a role in axonal guidance (Morgan, 2006)

as well as a suggested role in Parkinson's disease and autism (Gharani et al., 2004; Haubenberger et al., 2009; Morgan, 2006).

We could also show that the *MAMDC1* mRNA was expressed in the immune system and furthermore that it was up-regulated by the pro-inflammatory cytokines TNF-α and IFN-γ in THP-1 monocytes (Figure 7 in Paper II). These results suggest that *MAMDC1* expression could increase during inflammation and given its putative function as an adhesion molecule it is possible that this might have an impact on SLE disease predisposition and/or manifestations. This, of cause, raised the question if *MAMDC1* mRNA is up-regulated in SLE patients. Unfortunately, the only available source of mRNA from SLE patients was that from leukocytes immortalized by Epstein-Barr virus transformation (i.e. B-cells) and these showed little to no expression of *MAMDC1* mRNA. Thus, this question remains unanswered. Migration of leukocytes to sites of inflammation is crucial to the pathogenesis and development of inflammatory lesions in SLE and other autoimmune disorders (Norman and Hickey, 2005). Although the mechanisms underlying this leukocyte redistribution are still not fully understood, adhesion molecules such as those of the IgCAM superfamily have been strongly implicated in the recruitment of immune cells to sites of inflammation, and changes in their expression have been reported in RA, multiple sclerosis, TID and SLE (Sfikakis and Mavrikakis, 1999).

MAMDC1 was not reported among the top findings in any of the SLE GWA studies, however, given the moderate effect of this gene on SLE susceptibility this is not unexpected. Furthermore, the large amount of SNPs contained within the gene (n = 2317) as well as their location in several LD blocks (not shown) and the fact that only 7% or less of variation in this region was included in the SLE GWA studies (Illumina, Affymetrix) probably decreased the chances of identifying *MAMDC1*. It could be worth mentioning that a SNP in the *LRFN5* gene, located on chromosome14q21.1 and thus in our initial linkage region, was among the top 50 associations (*P*-value = 3.900 x 10^{-5}) in one of the GWA studies (Hom et al., 2008). However, when analyzing this SNP for association in our Finnish family cohort we did not find any significant association (data not shown), thus *LRFN5* does not appear to have contributed to our initial linkage peak. Yet, we have not investigated any of the other genes located within this region and thus additional susceptibility genes may still be harbored on the chromosome 14q21-q23 linkage region.

5.3 PAPER III, IV AND V – SEVERAL OF THE PREVIOUSLY IDENTIFIED SLE SUSCEPTIBLY GENES SHOWED ASSOCIATION ALSO IN FINNISH SLE PATIENTS

A large number of genes have been implicated in SLE susceptibility through positional mapping, functional candidate gene association studies and, most importantly, GWA studies (Table 2). Replication of previous association across both similar and diverse populations is of importance to validate these associations and also to get a better

understanding of potential population differences (Cunninghame Graham, 2009). In terms of GWA studies, associations that have been identified from a single data set rarely have definitive statistical support and a the required genome-wide significant *P*-value of $<10^{-7}$ only corresponds to a *P*-value of approximately 0.05 for a classical epidemiological study in which only one hypothesis is being tested (Ioannidis et al., 2009). Given the many analyses and outcomes that are assessed in a GWA study, even a *P*-value of 10^{-10} or less might be necessary to safely confirm an association (Pe'er et al., 2008). Furthermore, the Finnish population has been shown to differ from other Northern European populations and suggested to have a small eastern contribution (Salmela et al., 2008). Thus, in Papers III-V we wanted to investigate whether previously identified SLE susceptibility genes contribute to risk of SLE also in the Finnish population.

In Paper IV, which was done in collaboration with the group of Professor Ann-Christine Syvänen, the aim was to investigate if SNPs and haplotypes in *STAT4* previously associated with SLE in a Swedish case-control cohort (Sigurdsson et al., 2008) also were associated with increased risk in our Finnish family cohort. By TDT analysis we found the strongest signal of association for two linked SNPs: rs7582694 (*P*-value = 0.002, OR = 2.570) and rs10181656 (*P*-value = 0.001, OR = 2.530). We further performed haplotype association analysis using a sliding window approach which showed that the strongest association signal (*P*-value = 0.002, OR = 3.330) originated from SNPs in intron 3 of *STAT4*. This study had a partial overlap with Paper V, since rs3821236, rs10181656 and rs7582694 were genotyped in both studies, however, the approaches used for data analysis differed.

In Papers III and V, we used a combination of our two Finnish cohorts in order to obtain the maximal statistical power of association analysis and therefore one proband from each family was added to the SLE case-control cohort. In Paper III, which was initiated prior to the publication of the four GWA studies (Graham et al., 2008; Harley et al., 2008; Hom et al., 2008, Kozyrev et al., 2008), we investigated SNPs in seven genes suggested to be of importance in SLE susceptibility, including *TYK2*, *IRF5*, *CTLA4*, *PDCD1*, *FCGR2A*, *PTPN22* (failed in the genotyping) and *NOD2* (Table 2). Significant association was seen for rs2304256 (*P*-value = 0.0001, OR = 1.680) and rs12720270 (*P*-value = 0.003, OR = 1.570) in *TYK2* and rs10954213 (*P*-value = 0.004, OR = 1.420) in *IRF5*, but not in any the other genes (Table 2 in Paper III). In Paper V, which was initiated after the publication of the four GWA studies, the focus for our investigations was genes identified or confirmed in these studies (Table 2). We subsequently genotyped SNPs in 14 of the best supported GWA-identified SLE genes and loci, including *STAT4*, *BANK1*, *TNFAIP3*, *IRF5-TNPO3*, *FAM167A-BLK*, *KIAA1542*, *PXK*, *ITGAM-ITGAX*, *1q25.1*, *ATG5*, *ICA1*, *LYN*, *XKR6* and *SCUBE1* and could replicate findings for *STAT4*, *IRF5-TNPO3*, *ITGAM-ITGAX*, *TNFAIP3*, *FAM167A-BLK*, *BANK1* and *KIAA1542* (Table 2 in Paper V), but not for any of the other genes (Supplementary Table 3 in Paper V).

Significant haplotype associations were further identified in *TYK2* (Table 3 in Paper III), *STAT4*, *IRF5-TNPO3*, *ITGAM-ITGAX*, *TNFAIP3* and *FAM167A-BLK* (Table 3 in Paper V). However, this analysis did neither improve statistical significance nor provide additional information, since most of the detected haplotypic associations were due to individual SNPs and their alleles as from the previous single-marker analysis. This is not surprising, given the strong LD occurring at most of the studied loci, and the fact that a larger number of cases and controls would likely be needed to effectively test different haplotypes at each locus with sufficient power.

Out of the studied GWA-identified genes we were unable to replicate *1q25.1, PXK, ATG5, ICA1, XKR6, LYN* and *SCUBE1*. Both *PXK* and *1q25.1* have reached genome-wide level significance in a single GWA study and have been replicated in an independent case-control study (Suarez-Gestal et al., 2009), while the others have only reached suggestive GWA association (Table 1 in Paper V). However, the power to detect association in our sample for these genes was very modest (between 28% and 49%) and thus we cannot exclude them as susceptibility genes in the Finnish population.

Of the genes included in Papers III-V, *ITGAM-ITGAX*, *IRF5-TNPO3*, *FAM167A-BLK* and *STAT4* have reached genome-wide significance in at least two of the GWA studies and are considered as certain risk factors in SLE. *PTPN22* (included in Paper III but failed in the genotyping) and *FCGR2A* are considered as strong risk factors, given their high level of significance in one of the GWA studies and support from a meta-analysis of fine-mapping studies. A large number of genes show good evidence of association, i.e. they have showed a high level of significance in one of the GWA studies or in a fine-mapping study in at least two independent cohorts. Others, including *PDCD1*, *NOD2*, *TYK2* and *CTLA4-ICOS* still require further evidence of association before their role in SLE susceptibility can be fully understood, either because they have shown modest association, or because the literature is inconsistent (Table 2 and (Rhodes and Vyse, 2008)). The results obtained in Papers III, IV and V are in line with this, with some exceptions. *FCGR2A*, which is considered as a strong candidate in SLE susceptibility could not be replicated. However, lack of replication may have been a result of the modest power of our study to detect association to this gene (42%). On the other hand we found association to *TYK2*, which is a somewhat controversial susceptibility gene since association to this gene could not be confirmed in any of the GWA studies. Nevertheless, following the initial association of *TYK2* with SLE susceptibility (Sigurdsson et al., 2005) two studies in addition to ours have verified this result (Graham et al., 2007; Suarez-Gestal et al., 2009).

Very little evidence of gene-gene interaction (epistasis) has been found between SLE susceptibility genes. In the GWA studies this may be a consequence of the stringent thresholds set for multiple testing. Two studies have, however, shown additive effect in individuals carrying multiple risk alleles of *IRF5* and *STAT4* (Abelson et al., 2008; Sigurdsson et al., 2008), both members of the type 1 IFN pathway. Thus, in Paper III we wanted to investigate, if the studied SNP in *IRF5* possibly also show additive

effects with SNPs in *TYK2*, which is another member of this pathway, and found that the risk of SLE increased as a function of the number of risk factors (Table 3 in Paper III). Furthermore, significant overall interaction could be observed between rs10954213 in *IRF5* and rs2304256 in TYK2 (*P*-value = 0.014), where the risk alleles for rs10954213 in *IRF5* (AA) and rs2304256 in *TYK2* (CC) contributed most to the overall interaction (*P*-value < 0.0001, OR = 2.730; Figure 1 in Paper III). In Paper V we set out to test for possible gene-gene interaction for all SNPs included in the study, but could not detect any significant evidence of epistasis.

6 CONCLUDING REMARKS AND FURTHER PERSPECTIVE

There has been a rapid change in the field of genetic research during my time as a PhD student, especially by the introduction of GWA studies. When looking back at the contemporary literature from my first year as a PhD student (2003/2004) the genes implicated in SLE susceptibility were the *FcγR* genes, the *MHC*, *PDCD1*, *CRP*, *CTLA4*, *MBL* and a few others (Nath et al., 2004a; Tsao, 2004), while today the number of susceptibility genes is well over 30 (Table 2). Still, for the majority of complex diseases, including SLE, only a fraction of the inherited risk has been explained (Frazer et al., 2009) and thus several more susceptibility genes remain to be identified.

As a result of this thesis we have contributed with *GIMAP5* (Paper I) and *MAMDC1* (Paper II) to the list of genes implicated in SLE susceptibility. The finding of *MAMDC1* was especially rewarding since the course leading to this identification was initiated over 10 years ago and starting with the collection of Finnish SLE families suitable for linkage studies (Koskenmies et al., 2001). *MAMDC1* is also an intriguing finding given that so little is known about this gene. We have been able to shed some light on a potential role of *MAMDC1* in the pathogenesis of SLE, however, a lot of work remains before we will fully understand its exact role. Furthermore, we have not fully investigated the linkage region on chromosome 14q21-q23 and thus additional risk genes may be contained within this region. This is the case also for the two regions on chromosome 5p and 6q that, in addition to the region on chromosome 14q, showed suggestive linkage in the previous genome-wide linkage scan, with following fine-mapping (Koskenmies et al., 2004a; Koskenmies et al., 2004c). For *GIMAP5* the picture is a bit clearer since its role in apoptosis makes it a good candidate in the pathogenesis of SLE. However, solid genetic evidence for this association is still lacking given that only one additional study has shown association between this gene and SLE (Lim et al., 2009).

Associated genetic variants are not necessarily themselves functional variants. Finding these variants is vital in the understanding of the biological processes underlying disease pathogenesis in SLE and may provide information for developing new treatments. However, since identification of functional variants is not an easy task, it is important that the genes that are chosen for further studies have been replicated in several populations to really confirm their role in SLE susceptibility. Comprehensive replication of GWA studies is still lacking and in addition to our study (Paper V) only one such study has been accomplished to date (Suarez-Gestal et al., 2009). However, the modest size of the sample cohort in our replication studies (Papers III-V) precludes us from excluding non-replicated genes as true SLE susceptibility genes. Despite our modest power we were still able to replicate a number of genes and loci, including *STAT4*, *IRF5-TNPO3*, *TYK2*, *ITGAM-ITGAX*, *TNFAIP3*, *FAM167A-BLK*, *BANK1* and *KIAA1542* and furthermore show the presence of gene-gene interaction between SNPs in *IRF5* and *TYK2*.

7 ACKNOWLEGMENT

The work in this thesis was performed at the department of Bioscience and Nutrition at Karolinska Institutet, Stockholm, between the years 2003 and 2009. The journey of going through a PhD project is not one that I would have been able to do alone and I am truly grateful for the support I have been given from all of you during these years!

Först av allt vill jag tacka Juha Kere för att du gav mig möjligheten att få doktorera i din grupp. Du har verkligen lyckats att samla genuint trevliga och bra människor omkring dig och du har även en förmågan att inspirera som är beundransvärd. Trots att den senaste tiden har varit mycket svår för dig så har du ändå bibehållit din positiva attityd och varit till stort stöd i slutförandet av denna avhandling och för det är jag ytterst tacksam!

I would also like to thank my co-supervisor Mauro for all your support! The knowledge that you possess, as well as your passion for science and strive for perfection is remarkable and I have truly learnt a lot by having you at my side. Stort tack även till min andra bihandledare Cissi för du gav mig en plats i SLE projektet och tog hand om mig under mina första år som doktorand. Tack också för att du ville dela några soliga dagar med mig på Stora barriärrevet ☺. You have both been a great support to me in the different phases of this thesis work and I appreciate it!

Many thanks as well to everyone that has contributed in the SLE project. Tiina J, you have been a rock through this and I am so grateful for all the time you have spent proofreading my papers and thesis. Thank you also for introducing me to the world of farming just when I needed it ☺. Stort tack till Marco för all din hjälp med de statistiska analyserna, du har varit ett enormt stöd och jag är tacksam för dit stora tålamod. Thanks also to Sari, Heikki and Elisabeth for collecting the Finnish family cohort which has been the basis for this thesis and to Päivi for your great work in the identification of *MAMDC1*. Stort tack även till Tiina S och Ingegerd för all er hjälp och stöd i labbet, jag har lärt mig oerhört mycket utav er! Also I want to thank Alli for teaching me the immunohistochemistry method and Riitta for excellent technical assistance. Finally I want to thank Ulpu for being such a tremendous asset to this project and who I am very sad is no longer with us.

I would also want to give a huge thanks to all collaborators that have contributed samples and ideas to the different studies in this thesis! Thanks Tim, Debbie and Andy for your great hospitality when Cissi and I visited you in the UK and for all your help with the *GIMAP5* and *MAMDC1* papers. Tack till Elisabet, Iva och Leonid för ett givande och trevligt samarbete. Thanks also to Taina and Jaana for a nice collaboration. Tack även till Ann-Christine, Johanna och Sophie för ett fint samarbete i sambend med *STAT4* studien.

A special thanks to Vincenzo Casolaro for taking me on as a trainee prior to my PhD-studies and for introducing me into a world of T-cells, COX2-inhibitors and NSAIDs. I am grateful for your belief in me and for your encouragement!

Från MAF vill jag speciellt tacka Linda, Malin, Aida och Ville som alla haft en viktig roll i SLE projekt och även hjälpt mig i labbet! Tack också Kristina för att du har låtit mig köra mina projekt på MAF, det har varit mycket lärorikt! Självklart vill jag även tacka Astrid, Päivi, Eva, Lotta, Tiina och Ulf för att ni också alltid har ställt upp om jag har behövt hjälp. Det har varit ett nöje att dela min tid som doktorand med er på MAF och jag vill tacka för att ni har varit så bra rums-, lunch- och fikakompisar och för att ni har stöttat mig i både framgångar och motgångar.

Ett stort tack till nuvarande och gamla medlemmar i JKE; Juha, Mauro, Cissi, Myriam, Isabel, Kristiina, Ulf, Sara, Ingegerd, Marie, Hanna, Per, Gustaf, Cilla, Virpi, Heidi, Christina, Tiina, Hong, Marco, Liselott, Fransesca A, Hans, Fransesca B, Lovisa, Linda, Astrid, Aida, Malin, Eva, Lotta, Tiina, Ville, Kristina, Päivi, Thomas, Erik, Katja, Rasko, Pu och Erja. Tack till tjejerna i TMK, Astrid, Christina, Hanna, Heidi, Kristiina, Linda och Sara för supertrevliga och goda fredagsmiddagar med en för det mesta lagom dos av vin och skvaller ☺. Speciellt tack till Linda och Hanna för att ni har kommit att betyda så mycket för mig och har varit så fantastiska vänner under de här åren!

Tack även till alla i dojon för roliga träningar och trevliga karate-öl kvällar!

Stort tack till mina vänner från Med Bi: Anna, Louise, Johanna L, Angelica, Calle, Kalle, Henrik, Sebastian och Anton för att ni bidrog till att göra åren i Linköping till en superbra och rolig tid! Anna, jag är tacksam att ha någon som du som vän och för allt kul vi har haft ihop genom åren. Louise, för att du är en så genuint snäll och omtänksam person! Tack Johanna L för trevliga luncher i Solna och Calle för trevliga promenader och boxningsmatcher. Johanna A, tack för att du hamnade i mig grupp på MOS-kursen och har varit en så bra vän ända sen dess ☺.

Även ett stort tack till alla mina vänner från Örebro! Malin, det betyder så mycket för mig att vi efter så många år på olika håll fortfarande kan umgås som att vi sågs varje dag! Pernilla, tack för några roliga veckor i Italien ☺. Tack även Susanne, Linda, Marie, Camilla, Carro, Malin P och Sara för trevliga fikastunder, kortspelskvällar och en väldigt lyckad 30-års fest!

Tack till alla nya vänner här i Stockholm som har förgyllt min fritid: Johanna S, Monika, Ying, Anna A, John, Jenny, Lina, Mattias och toastmaster Jenny!

Jag vill även tacka mina släktingar Ulla, Berton, Erik, Elin, Marianne, Stefan, Henrik, Kristian, Olof och farmor Signe för trevliga kalas och ert intresse för min forskning! Speciellt tack till farmor Signe för bonusresan, det var riktigt spännande att få se dina hembyggder!

Ett jättestort tack till mina föräldrar för allt ert stöd och uppmuntran genom åren! Att jag började doktorera, med allt vad det innebär, är ingen slump utan en konsekvens av ert förtroende i att jag kan genomföra det jag åtagit mig. Det uppskattar jag! Även tack pappa Ingvar för all fantastisk hjälp i samband med renoveringen av min lägenhet och tack mamma Lena för trevligt sällskap på välbehövliga vårresor runt om i Europa! Tack lillebror Rikard för att du är den du är och även för all din hjälp i samband med flytten! Du vet väl att du måste kalla mig Doktor Anna från och med nu och för all framtid ☺.

Sist men inte minst vill jag tacka Torbjörn ♥ för att du är så fantastisk! Det finns inte ord för hur mycket jag uppskattar all din hjälp under de här senaste månaderna när jag har suttit uppslukad i avhandlingen! Du är det bästa som hänt mig och jag är så glad för att jag träffat dig ♥.

8 REFERENCES

Abelson, A.K., Delgado-Vega, A.M., Kozyrev, S.V., Sanchez, E., *et al.* (2008). STAT4 Associates with SLE through two independent effects that correlate with gene expression and act additively with IRF5 to increase risk. Annals of the rheumatic diseases Dec 9, [Epub ahead of print].

Aitman, T.J., Dong, R., Vyse, T.J., Norsworthy, P.J., *et al.* (2006). Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature 439:851-855.

Alarcon-Segovia, D., Alarcon-Riquelme, M.E., Cardiel, M.H., Caeiro, F., *et al.* (2005). Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. Arthritis Rheum 52:1138-1147.

Altshuler, D., Daly, M.J., and Lander, E.S. (2008). Genetic mapping in human disease. Science 322:881-888.

Alwine, J.C., Kemp, D.J., and Stark, G.R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc Natl Acad Sci U S A 74:5350-5354.

Arbuckle, M.R., McClain, M.T., Rubertone, M.V., Scofield, R.H., *et al.* (2003). Development of autoantibodies before the clinical onset of systemic lupus erythematosus. N Engl J Med 349:1526-1533.

Aringer, M., and Smolen, J.S. (2008). The role of tumor necrosis factor-alpha in systemic lupus erythematosus. Arthritis research & therapy 10:202.

Armstrong, D.L., Reiff, A., Myones, B.L., Quismorio, F.P., *et al.* (2009). Identification of new SLE-associated genes with a two-step Bayesian study design. Genes Immun 10:446-456.

Balow, J.E. (2005). Clinical presentation and monitoring of lupus nephritis. Lupus 14:25-30.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics (Oxford, England) 21:263-265.

Bell, P.A., Chaturvedi, S., Gelfand, C.A., Huang, C.Y., *et al.* (2002). SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. BioTechniques *Suppl*, 70-72, 74, 76-77.

Bengtsson, A.A., Sturfelt, G., Truedsson, L., Blomberg, J., *et al.* (2000). Activation of type I interferon system in systemic lupus erythematosus correlates with disease activity but not with antiretroviral antibodies. Lupus 9:664-671.

Boehnke, M., and Langefeld, C.D. (1998). Genetic association mapping based on discordant sib pairs: the discordant-alleles test. Am J Hum Genet 62:950-961.

Borecki, I.B., and Province, M.A. (2008). Linkage and association: basic concepts. Advances in genetics 60:51-74.

Brandes, M., and Bienvenu, O.J. (2006). Personality and anxiety disorders. Current psychiatry reports 8:263-269.

Brown, E.E., Edberg, J.C., and Kimberly, R.P. (2007). Fc receptor genes and the systemic lupus erythematosus diathesis. Autoimmunity 40:567-581.

Bucan, M., Abrahams, B.S., Wang, K., Glessner, J.T., *et al.* (2009). Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. PLoS genetics 5:e1000536.

Cantor, R.M., Yuan, J., Napier, S., Kono, N., *et al.* (2004). Systemic lupus erythematosus genome scan: support for linkage at 1q23, 2q33, 16q12-13, and 17q21-23 and novel evidence at 3p24, 10q23-24, 13q32, and 18q22-23. Arthritis Rheum 50:3203-3210.

Cardon, L.R., and Bell, J.I. (2001). Association study designs for complex diseases. Nature reviews 2:91-99.

Cohen, P.L. (2006). Apoptotic cell death and lupus. Springer seminars in immunopathology 28:145-152.

Cordell, H.J. (2002). Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet 11:2463-2468.

Costenbader, K.H., and Karlson, E.W. (2005). Cigarette smoking and systemic lupus erythematosus: a smoking gun? Autoimmunity 38:541-547.

Coxon, A., Rieu, P., Barkalow, F.J., Askari, S., *et al.* (1996). A novel role for the beta 2 integrin CD11b/CD18 in neutrophil apoptosis: a homeostatic mechanism in inflammation. Immunity 5:653-666.

Criswell, L.A. (2008). The genetic contribution to systemic lupus erythematosus. Bulletin of the NYU hospital for joint diseases 66:176-183.

Crow, M.K., and Kirou, K.A. (2004). Interferon-alpha in systemic lupus erythematosus. Current opinion in rheumatology 16:541-547.

Cunninghame Graham, D.S. (2009). Genome-wide association studies in systemic lupus erythematosus: a perspective. Arthritis research & therapy 11:119.

Cunninghame Graham, D.S., Graham, R.R., Manku, H., Wong, A.K., *et al.* (2008). Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus. Nat Genet 40:83-89.

Cunninghame Graham, D.S., Manku, H., Wagner, S., Reid, J., *et al.* (2007). Association of IRF5 in UK SLE families identifies a variant involved in polyadenylation. Hum Mol Genet 16:579-591.

Daly, M., Rioux, J.D., Schaffner S.F., Hudson T.J., and Lander E.S. (2001). High-resolution haplotype structure in the human genome. Nat Genet 29:229-32

Danchenko, N., Satia, J.A., and Anthony, M.S. (2006). Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. Lupus *15*, 308-318.

David, M. (2002). Signal transduction by type I interferons. BioTechniques *Suppl*, 58-65.

Deapen, D., Escalante, A., Weinrib, L., Horwitz, D., *et al.* (1992). A revised estimate of twin concordance in systemic lupus erythematosus. Arthritis Rheum 35:311-318.

Decker, T., Stockinger, S., Karaghiosoff, M., Muller, M., and Kovarik, P. (2002). IFNs and STATs in innate immunity to microorganisms. The Journal of clinical investigation 109:1271-1277.

Demas, K.L., and Costenbader, K.H. (2009). Disparities in lupus care and outcomes. Current opinion in rheumatology 21:102-109.

Diveu, C., McGeachy, M.J., and Cua, D.J. (2008). Cytokines that regulate autoimmunity. Current opinion in immunology 20:663-668.

Dudbridge, F. (2003). Pedigree disequilibrium tests for multilocus haplotypes. Genetic epidemiology 25:115-121.

Dudbridge, F. (2008). Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. Human heredity 66:87-98.

Edberg, J.C., Langefeld, C.D., Wu, J., Moser, K.L., *et al.* (2002). Genetic linkage and association of Fcgamma receptor IIIA (CD16A) on chromosome 1q23 with human systemic lupus erythematosus. Arthritis Rheum 46:2132-2140.

Edberg, J.C., Wu, J., Langefeld, C.D., Brown, E.E., *et al.* (2008). Genetic variation in the CRP promoter: association with systemic lupus erythematosus. Hum Mol Genet 17:1147-1155.

Edwalds-Gilbert, G., Veraldi, K.L., and Milcarek, C. (1997). Alternative poly(A) site selection in complex transcription units: means to an end? Nucleic acids research *25*:2547-2561.

Eisenberg, R. (2009). Why can't we find a new treatment for SLE? Journal of autoimmunity 32:223-230.

Esdaile, J.M., Abrahamowicz, M., Grodzicky, T., Li, Y., *et al.* (2001). Traditional Framingham risk factors fail to fully account for accelerated atherosclerosis in systemic lupus erythematosus. Arthritis Rheum 44:2331-2337.

Fanciulli, M., Norsworthy, P.J., Petretto, E., Dong, R., *et al.* (2007). FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. Nat Genet 39:721-723.

Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. Nature reviews 7:85-97.

Fielder, A.H., Walport, M.J., Batchelor, J.R., Rynes, R.I., *et al.* (1983). Family study of the major histocompatibility complex in patients with systemic lupus erythematosus: importance of null alleles of C4A and C4B in determining disease susceptibility. British medical journal (Clinical research ed) 286:425-428.

Forabosco, P., Gorman, J.D., Cleveland, C., Kelly, J.A., Fisher, S.A., Ortmann, W.A., Johansson, C., Johanneson, B., Moser, K.L., Gaffney, P.M.*, et al.* (2006). Meta-analysis of genome-wide linkage studies of systemic lupus erythematosus. Genes Immun 7:609-614.

Frazer, K.A., Murray, S.S., Schork, N.J., and Topol, E.J. (2009). Human genetic variation and its contribution to complex traits. Nature reviews 10:241-251.

Gaffney, P.M., Kearns, G.M., Shark, K.B., Ortmann, W.A., *et al.* (1998). A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. Proc Natl Acad Sci U S A 95:14875-14879.

Gaffney, P.M., Ortmann, W.A., Selby, S.A., Shark, K.B., *et al.* (2000). Genome screening in human systemic lupus erythematosus: results from a second Minnesota cohort and combined analyses of 187 sib-pair families. Am J Hum Genet 66:547-556.

Garred, P., Voss, A., Madsen, H.O., and Junker, P. (2001). Association of mannose-binding lectin gene variation with disease severity and infections in a population-based cohort of systemic lupus erythematosus patients. Genes Immun 2:442-450.

Gharani, N., Benayed, R., Mancuso, V., Brzustowicz, L.M., and Millonig, J.H. (2004). Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder. Molecular psychiatry 9:474-484.

Graham, D.S., Akil, M., and Vyse, T.J. (2007). Association of polymorphisms across the tyrosine kinase gene, TYK2 in UK SLE families. Rheumatology (Oxford) 46:927-930.

Graham, D.S., Wong, A.K., McHugh, N.J., Whittaker, J.C., and Vyse, T.J. (2006). Evidence for unique association signals in SLE at the CD28-CTLA4-ICOS locus in a family-based study. Hum Mol Genet 15:3195-3205.

Graham, R.R., Cotsapas, C., Davies, L., Hackett, R., *et al.* (2008). Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nat Genet. 40:1059-61

Gray-McGuire, C., Moser, K.L., Gaffney, P.M., Kelly, J., *et al.* (2000). Genome scan of human systemic lupus erythematosus by regression modeling: evidence of linkage and epistasis at 4p16-15.2. Am J Hum Genet 67:1460-1469.

Gregersen, P.K., and Olsson, L.M. (2009). Recent advances in the genetics of autoimmune disease. Annu Rev Immunol 27:363-391.

Grondal, G., Gunnarsson, I., Ronnelid, J., Rogberg, S., Klareskog, L., and Lundberg, I. (2000). Cytokine production, serum levels and disease activity in systemic lupus erythematosus. Clinical and experimental rheumatology 18:565-570.

Han, S., Guthridge, J.M., Harley, I.T., Sestak, A.L., *et al.* (2008). Osteopontin and systemic lupus erythematosus association: a probable gene-gender interaction. PloS one 3:e0001757.

Hardy, M.P., Owczarek, C.M., Jermiin, L.S., Ejdeback, M., and Hertzog, P.J. (2004). Characterization of the type I interferon locus and identification of novel genes. Genomics 84:331-345.

Harley, I.T., Kaufman, K.M., Langefeld, C.D., Harley, J.B., and Kelly, J.A. (2009). Genetic susceptibility to SLE: new insights from fine mapping and genome-wide association studies. Nature reviews 10:285-290.

Harley, J.B., Alarcon-Riquelme, M.E., Criswell, L.A., Jacob, C.O., *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.

Harley, J.B., Harley, I.T., Guthridge, J.M., and James, J.A. (2006a). The curiously suspicious: a role for Epstein-Barr virus in lupus. Lupus 15:768-777.

Harley, J.B., Kelly, J.A., and Kaufman, K.M. (2006b). Unraveling the genetics of systemic lupus erythematosus. Springer seminars in immunopathology 28:119-130.

Haubenberger, D., Reinthaler, E., Mueller, J.C., Pirker, W., *et al.* (2009). Association of transcription factor polymorphisms PITX3 and EN1 with Parkinson's disease. Neurobiology of aging Apr 2, [Epub ahead of print]

Hellquist, A., Zucchelli, M., Kivinen, K., Saarialho-Kere, U., *et al.* (2007). The human GIMAP5 gene has a common polyadenylation polymorphism increasing risk to systemic lupus erythematosus. J Med Genet 44:314-321.

Helmick, C.G., Felson, D.T., Lawrence, R.C., Gabriel, S., *et al.* (2008). Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. Arthritis Rheum 58:15-25.

Helve, T. (1985). Prevalence and mortality rates of systemic lupus erythematosus and causes of death in SLE patients in Finland. Scandinavian journal of rheumatology 14:43-46.

Hirschhorn, J.N., and Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. Nature reviews 6:95-108.

Hirschhorn, J.N., Lohmueller, K., Byrne, E., and Hirschhorn, K. (2002). A comprehensive review of genetic association studies. Genet Med 4:45-61.

Hochberg, M.C. (1987). The application of genetic epidemiology to systemic lupus erythematosus. J Rheumatol 14:867-869.

Hochberg, M.C. (1997). Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 40:1725.

Hom, G., Graham, R.R., Modrek, B., Taylor, K.E., *et al.* (2008). Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med 358:900-909.

Honda, K., and Taniguchi, T. (2006). IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. Nat Rev Immunol 6:644-658.

Hornum, L., Romer, J., and Markholst, H. (2002). The diabetes-prone BB rat carries a frameshift mutation in Ian4, a positional candidate of Iddm1. Diabetes 51:1972-1979.

Hu, G., and Barnes, B.J. (2009). IRF-5 is a mediator of the death receptor-induced apoptotic signaling pathway. J Biol Chem 284:2767-2777.

Ioannidis, J.P., Thomas, G., and Daly, M.J. (2009). Validating, augmenting and refining genome-wide association signals. Nature reviews 10:318-329.

Isenberg, D.A., and Collins, C. (1985). Detection of cross-reactive anti-DNA antibody idiotypes on renal tissue-bound immunoglobulins from lupus patients. The Journal of clinical investigation 76:287-294.

Isenberg, D.A., Manson, J.J., Ehrenstein, M.R., and Rahman, A. (2007). Fifty years of anti-ds DNA antibodies: are we approaching journey's end? Rheumatology (Oxford) 46:1052-1056.

Jacob, C.O., Reiff, A., Armstrong, D.L., Myones, *et al.* (2007). Identification of novel susceptibility genes in childhood-onset systemic lupus erythematosus using a uniquely designed candidate gene pathway platform. Arthritis Rheum 56:4164-4173.

Jacob, C.O., Zhu, J., Armstrong, D.L., Yan, M., *et al.* (2009). Identification of IRAK1 as a risk gene with critical role in the pathogenesis of systemic lupus erythematosus. Proc Natl Acad Sci U S A 106:6256-6261.

Jailwala, P., Waukau, J., Glisic, S., Jana, S., *et al.* (2009). Apoptosis of CD4+ CD25(high) T cells in type 1 diabetes may be partially mediated by IL-2 deprivation. PloS one 4:e6527.

James, J.A., Neas, B.R., Moser, K.L., Hall, T., *et al.* (2001). Systemic lupus erythematosus in adults is associated with previous Epstein-Barr virus exposure. Arthritis Rheum 44:1122-1126.

Janko, C., Schorn, C., Grossmayer, G.E., Frey, B., *et al.* (2008). Inflammatory clearance of apoptotic remnants in systemic lupus erythematosus (SLE). Autoimmunity reviews 8:9-12.

Johansson, C.M., Zunec, R., Garcia, M.A., Scherbarth, *et al.* (2004). Chromosome 17p12-q11 harbors susceptibility loci for systemic lupus erythematosus. Hum Genet 115:230-238.

Johnson, A.E., Gordon, C., Palmer, R.G., and Bacon, P.A. (1995). The prevalence and incidence of systemic lupus erythematosus in Birmingham, England. Relationship to ethnicity and country of birth. Arthritis Rheum 38:551-558.

Jurinke, C., van den Boom, D., Cantor, C.R., and Koster, H. (2002). Automated genotyping using the DNA MassArray technology. Methods Mol Biol 187:179-192.

Kahler, A.K., Djurovic, S., Kulle, B., Jonsson, E.G., *et al.* (2008). Association analysis of schizophrenia on 18 genes involved in neuronal migration: MDGA1 as a new susceptibility gene. Am J Med Genet B Neuropsychiatr Genet 147B:1089-1100.

Kaplan, M.H. (2005). STAT4: a critical regulator of inflammation in vivo. Immunologic research 31:231-242.

Karassa, F.B., Trikalinos, T.A., and Ioannidis, J.P. (2002). Role of the Fcgamma receptor IIa polymorphism in susceptibility to systemic lupus erythematosus and lupus nephritis: a meta-analysis. Arthritis Rheum 46:1563-1571.

Kariuki, S.N., Moore, J.G., Kirou, K.A., Crow, M.K., *et al.* (2009). Age- and gender-specific modulation of serum osteopontin and interferon-alpha by osteopontin genotype in systemic lupus erythematosus. Genes Immun 10:487-494.

Korman, B.D., Kastner, D.L., Gregersen, P.K., and Remmers, E.F. (2008). STAT4: genetics, mechanisms, and implications for autoimmunity. Current allergy and asthma reports 8:398-403.

Koskenmies, S., Jarvinen, T.M., Onkamo, P., Panelius, J., *et al.*. (2008). Clinical and laboratory characteristics of Finnish lupus erythematosus patients with cutaneous manifestations. Lupus 17:337-347.

Koskenmies, S., Lahermo, P., Julkunen, H., Ollikainen, V., *et al.* (2004a). Linkage mapping of systemic lupus erythematosus (SLE) in Finnish families multiply affected by SLE. J Med Genet 41:e2-5

Koskenmies, S., Vaarala, O., Widen, E., Kere, J., *et al.*. (2004b). The association of antibodies to cardiolipin, beta 2-glycoprotein I, prothrombin, and oxidized low-density lipoprotein with thrombosis in 292 patients with familial and sporadic systemic lupus erythematosus. Scandinavian journal of rheumatology 33:246-252.

Koskenmies, S., Widen, E., Kere, J., and Julkunen, H. (2001). Familial systemic lupus erythematosus in Finland. J Rheumatol 28:758-760.

Koskenmies, S., Widen, E., Onkamo, P., Sevon, P., *et al.* (2004c). Haplotype associations define target regions for susceptibility loci in systemic lupus erythematosus. Eur J Hum Genet 12, 489-494.

Kozyrev, S.V., Abelson, A.K., Wojcik, J., Zaghlool, A., *et al. et al.* (2008). Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nat Genet 40:211-216.

Kozyrev, S.V., and Alarcon-Riquelme, M.E. (2007). The genetics and biology of Irf5-mediated signaling in lupus. Autoimmunity 40:591-601.

Kruglyak, L., Daly, M.J., Reeve-Daly, M.P., and Lander, E.S. (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347-1363.

Kyogoku, C., Langefeld, C.D., Ortmann, W.A., Lee, A., *et al.* (2004). Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. Am J Hum Genet 75:504-507.

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

Laird, N.M., and Lange, C. (2008). Family-based methods for linkage and association analysis. Advances in genetics 60:219-252.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, *et al.* (2001). Initial sequencing and analysis of the human genome. Nature 409: 860-921.

Lander, E.S., and Schork, N.J. (1994). Genetic dissection of complex traits. Science (New York, N.Y *265*, 2037-2048.

Lau, C.S., Yin, G., and Mok, M.Y. (2006). Ethnic and geographical differences in systemic lupus erythematosus: an overview. Lupus 15:715-719.

Lee-Kirsch, M.A., Gong, M., Chowdhury, D., *et al.* (2007). Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 are associated with systemic lupus erythematosus. Nat Genet 39:1065-1067.

Lee, Y.H., Harley, J.B., and Nath, S.K. (2005). CTLA-4 polymorphisms and systemic lupus erythematosus (SLE): a meta-analysis. Hum Genet 116:361-367.

Lee, Y.H., and Nath, S.K. (2005). Systemic lupus erythematosus susceptibility loci defined by genome scan meta-analysis. Hum Genet 118:434-443.

Li, X., Ptacek, T.S., Brown, E.E., and Edberg, J.C. (2009). Fcgamma receptors: structure, function and role as genetic risk factors in SLE. Genes Immun 10:380-389.

Lim, M., Sheen, D., Kim, S., Won, S., *et al.* (2009). IAN5 polymorphisms are associated with systemic lupus erythematosus. Lupus 18:1045-1052.

Lindqvist, A.K., Steinsson, K., Johanneson, B., Kristjansdottir, H., *et al.* (2000). A susceptibility locus for human systemic lupus erythematosus (hSLE1) on chromosome 2q. Journal of autoimmunity *14*:169-178.

Linker-Israeli, M., Deans, R.J., Wallace, D.J., Prehn, J., *et al.* (1991). Elevated levels of endogenous IL-6 in systemic lupus erythematosus. A putative role in pathogenesis. J Immunol 147:117-123.

Litwack, E.D., Babey, R., Buser, R., Gesemann, M., and O'Leary, D.D. (2004). Identification and characterization of two novel brain-derived immunoglobulin

superfamily members with a unique structural organization. Molecular and cellular neurosciences 25:263-274.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402-408.

Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S., and Hirschhorn, J.N. (2003). Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33:177-182.

MacMurray, A.J., Moralejo, D.H., Kwitek, A.E., Rutledge, E.A., *et al.* (2002). Lymphopenia in the BB rat model of type 1 diabetes is due to a mutation in a novel immune-associated nucleotide (Ian)-related gene. Genome research 12:1029-1039.

Mallavarapu, R.K., and Grimsley, E.W. (2007). The history of lupus erythematosus. South Med J 100:896-898.

Manzi, S. (2009). Lupus update: perspective and clinical pearls. Cleveland Clinic journal of medicine 76:137-142.

Martin, E.R., Monks, S.A., Warren, L.L., and Kaplan, N.L. (2000). A test for linkage and association in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet 67:146-154.

Masi, A.T., and Kaslow, R.A. (1978). Sex effects in systemic lupus erythematosus: a clue to pathogenesis. Arthritis Rheum 21:480-484.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., *et al.* (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature reviews 9:356-369.

Merrell, M., and Shulman, L.E. (1955). Determination of prognosis in chronic disease, illustrated by systemic lupus erythematosus. J Chronic Dis 1:12-32.

Mok, C.C., and Lau, C.S. (2003). Pathogenesis of systemic lupus erythematosus. Journal of clinical pathology 56:481-490.

Molina, V., and Shoenfeld, Y. (2005). Infection, vaccines and other environmental triggers of autoimmunity. Autoimmunity 38:235-245.

Molokhia, M., and McKeigue, P. (2006). Systemic lupus erythematosus: genes versus environment in high risk populations. Lupus 15:827-832.

Moore, J.H., and Williams, S.M. (2005). Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. Bioessays 27:637-646.

Morgan, R. (2006). Engrailed: complexity and economy of a multi-functional transcription factor. FEBS letters 580:2531-2533.

Morris, D.L., Graham, R.R., Erwig, L.P., Gaffney, P.M., *et al.* (2009). Variation in the upstream region of P-Selectin (SELP) is a risk factor for SLE. Genes Immun 10:404-413.

Morton, N.E. (1955). Sequential tests for the detection of linkage. Am J Hum Genet 7:277-318.

Moser, K.L., Kelly, J.A., Lessard, C.J., and Harley, J.B. (2009). Recent insights into the genetic basis of systemic lupus erythematosus. Genes Immun 10:373-379.

Moser, K.L., Neas, B.R., Salmon, J.E., Yu, H., *et al.* (1998). Genome scan of human systemic lupus erythematosus: evidence for linkage on chromosome 1q in African-American pedigrees. Proceedings of the National Academy of Sciences of the United States of America 95:14869-14874.

Munoz, L.E., van Bavel, C., Franz, S., Berden, J., *et al.* (2008). Apoptosis in the pathogenesis of systemic lupus erythematosus. Lupus 17:371-375.

Musone, S.L., Taylor, K.E., Lu, T.T., Nititham, J., *et al.* (2008). Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. Nat Genet 40:1062-1064.

Narain, S., Richards, H.B., Satoh, M., Sarmiento, M., *et al.* 2004). Diagnostic accuracy for lupus and other systemic autoimmune diseases in the community setting. Archives of internal medicine 164:2435-2441.

Nath, S.K., Han, S., Kim-Howard, X., Kelly, J.A., *et al.* (2008). A nonsynonymous functional variant in integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus. Nat Genet 40:152-154.

Nath, S.K., Kilpatrick, J., and Harley, J.B. (2004a). Genetics of human systemic lupus erythematosus: the emerging picture. Current opinion in immunology 16:794-800.

Nath, S.K., Quintero-Del-Rio, A.I., Kilpatrick, J., Feo, L., *et al.* (2004b). Linkage at 12q24 with systemic lupus erythematosus (SLE) is established and confirmed in Hispanic and European American families. Am J Hum Genet 74:73-82.

Nicodemus, K.K. (2008). Catmap: case-control and TDT meta-analysis package. BMC bioinformatics 9:130.

Niewold, T.B., Hua, J., Lehman, T.J., Harley, J.B., and Crow, M.K. (2007). High serum IFN-alpha activity is a heritable risk factor for systemic lupus erythematosus. Genes Immun 8:492-502.

Norman, M.U., and Hickey, M.J. (2005). Mechanisms of lymphocyte migration in autoimmune disease. Tissue Antigens 66:163-172.

Nyholt, D.R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet 74:765-769.

O'Connell, J.R., and Weeks, D.E. (1998). PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259-266.

Olson, J.M., Song, Y., Dudek, D.M., Moser, K.L., *et al.* (2002). A genome screen of systemic lupus erythematosus using affected-relative-pair linkage analysis with covariates demonstrates genetic heterogeneity. Genes Immun 3:Suppl 1, S5-S12.

Orr, N., and Chanock, S. (2008). Common genetic variation and human disease. Advances in genetics 62:1-32.

Pandarpurkar, M., Wilson-Fritch, L., Corvera, S., Markholst, H., *et al.* (2003). Ian4 is required for mitochondrial integrity and T cell survival. Proc Natl Acad Sci U S A 100:10382-10387.

Pastinen, T., Sladek, R., Gurd, S., Sammak, A., *et al.* (2004). A survey of genetic and epigenetic variation affecting human gene expression. Physiological genomics 16:184-193.

Pe'er, I., Yelensky, R., Altshuler, D., and Daly, M.J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genetic epidemiology 32:381-385.

Petri, M. (2005). Review of classification criteria for systemic lupus erythematosus. Rheumatic diseases clinics of North America 31:245-254.

Petri, M. (2008). Sex hormones and systemic lupus erythematosus. Lupus *17*, 412-415.

Pino, S.C., O'Sullivan-Murphy, B., Lidstone, E.A., Yang, C., *et al.* (2009). CHOP mediates endoplasmic reticulum stress-induced apoptosis in Gimap5-deficient T cells. PloS one 4:e5468.

Priori, R., Medda, E., Conti, F., Cassara, E.A., *et al.* (2003). Familial autoimmunity as a risk factor for systemic lupus erythematosus and vice versa: a case-control study. Lupus 12:735-740.

Prokunina, L., Castillejo-Lopez, C., Oberg, F., Gunnarsson, I., *et al.* (2002). A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. Nat Genet 32:666-669.

Rahman, A., and Isenberg, D.A. (2008). Systemic lupus erythematosus. N Engl J Med 358:929-939.

Ramanujam, M., and Davidson, A. (2008). Targeting of the immune system in systemic lupus erythematosus. Expert reviews in molecular medicine 10:e2.

Remmers, E.F., Plenge, R.M., Lee, A.T., Graham, R.R., *et al.* (2007). STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 357:977-986.

Rhodes, B., and Vyse, T.J. (2007). General aspects of the genetics of SLE. Autoimmunity 40:550-559.

Rhodes, B., and Vyse, T.J. (2008). The genetics of SLE: an update in the light of genome-wide association studies. Rheumatology (Oxford) 47:1603-11

Richter, M.F., Dumenil, G., Uze, G., Fellous, M., and Pellegrini, S. (1998). Specific contribution of Tyk2 JH regions to the binding and the expression of the interferon alpha/beta receptor component IFNAR1. J Biol Chem 273:24723-24729.

Risch, N. (1990). Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222-228.

Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. Science (New York, N.Y *273*, 1516-1517.

Risch, N.J. (2000). Searching for genetic determinants in the new millennium. Nature 405:847-856.

Ronnblom, L., and Pascual, V. (2008). The innate immune system in SLE: type I interferons and dendritic cells. Lupus 17:394-399.

Russell, A.I., Cunninghame Graham, D.S., Shepherd, C., Roberton, C.A., *et al.* (2004). Polymorphism at the C-reactive protein locus influences gene expression and predisposes to systemic lupus erythematosus. Hum Mol Genet 13:137-147.

Salmela, E., Lappalainen, T., Fransson, I., Andersen, P.M., *et al.* (2008). Genome-wide analysis of single nucleotide polymorphisms uncovers population structure in Northern Europe. PloS one 3:e3519.

Sarzi-Puttini, P., Atzeni, F., Iaccarino, L., and Doria, A. (2005). Environment and systemic lupus erythematosus: an overview. Autoimmunity *38*, 465-472.

Sawalha, A.H., Webb, R., Han, S., Kelly, J.A., *et al.* (2008). Common variants within MECP2 confer risk of systemic lupus erythematosus. PloS one 3:e1727.

Schindler, C., and Plumlee, C. (2008). Inteferons pen the JAK-STAT pathway. Seminars in cell & developmental biology 19:311-8.

Schulteis, R.D., Chu, H., Dai, X., Chen, Y., *et al.* (2008). Impaired survival of peripheral T cells, disrupted NK/NKT cell development, and liver failure in mice lacking Gimap5. Blood 112:4905-4914.

Scofield, R.H., Bruner, G.R., Namjou, B., Kimberly, R.P., *et al.* (2008). Klinefelter's syndrome (47,XXY) in male systemic lupus erythematosus patients: support for the notion of a gene-dose effect from the X chromosome. Arthritis Rheum 58:2511-2517.

Sevon, P., Ollikainen, V., Onkamo, P., Toivonen, H.T., *et al.* (2001). Mining associations between genetic markers, phenotypes, and covariates. Genetic epidemiology 21: Suppl 1, S588-593.

Sfikakis, P.P., and Mavrikakis, M. (1999). Adhesion and lymphocyte costimulatory molecules in systemic rheumatic diseases. Clinical rheumatology 18:317-327.

Shai, R., Quismorio, F.P., Jr., Li, L., Kwon, O.J., *et al.* (1999). Genome-wide screen for systemic lupus erythematosus susceptibility genes in multiplex families. Hum Mol Genet 8:639-644.

Shin, J.H., Janer, M., McNeney, B., Blay, S., *et al.* (2007). IA-2 autoantibodies in incident type I diabetes patients are associated with a polyadenylation signal polymorphism in GIMAP5. Genes Immun 8:503-512.

Sigurdsson, S., Nordmark, G., Garnier, S., Grundberg, E., *et al.* (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.

Sigurdsson, S., Nordmark, G., Goring, H.H., Lindroos, K., *et al.* (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.

Simard, J.F., and Costenbader, K.H. (2007). What can epidemiology tell us about systemic lupus erythematosus? International journal of clinical practice 61:1170-1180.

Sousa, E., and Isenberg, D. (2009). Treating lupus: from serendipity to sense, the rise of the new biologicals and other emerging therapies. Best practice & research 23:563-574.

Spielman, R.S., McGinnis, R.E., and Ewens, W.J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506-516.

Stetson, D.B., Ko, J.S., Heidmann, T., and Medzhitov, R. (2008). Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell 134:587-598.

Suarez-Gestal, M., Calaza, M., Endreffy, E., Pullmann, R., *et al.* (2009). Replication of recently identified systemic lupus erythematosus genetic associations: a case-control study. Arthritis research & therapy 11:R69.

Sule, S., and Petri, M. (2006). Socioeconomic status in systemic lupus erythematosus. Lupus 15:720-723.

Sullivan, K.E., Petri, M.A., Schmeckpeper, B.J., McLean, R.H., and Winkelstein, J.A. (1994). Prevalence of a mutation causing C2 deficiency in systemic lupus erythematosus. J Rheumatol 21:1128-1133.

Svenungsson, E., Gunnarsson, I., Fei, G.Z., Lundberg, I.E., *et al.* (2003). Elevated triglycerides and low levels of high-density lipoprotein as markers of disease activity in association with up-regulation of the tumor necrosis factor alpha/tumor necrosis factor receptor system in systemic lupus erythematosus. Arthritis Rheum 48:2533-2540.

Syvanen, A.C. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. Nature reviews 2:930-942.

Tan, E.M., Cohen, A.S., Fries, J.F., Masi, A.T., *et al.* (1982). The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 25:1271-1277.

Taylor, K.E., Remmers, E.F., Lee, A.T., Ortmann, *et al.* (2008). Specificity of the STAT4 genetic association for severe disease manifestations of systemic lupus erythematosus. PLoS genetics 4:e1000084.

Toivonen, H.T., Onkamo, P., Vasko, K., Ollikainen, V., *et al.* (2000). Data mining applied to linkage disequilibrium mapping. Am J Hum Genet *67*, 133-145.

Tsao, B.P. (2004). Update on human systemic lupus erythematosus genetics. Current opinion in rheumatology 16:513-521.

Wain, L.V., Armour, J.A., and Tobin, M.D. (2009). Genomic copy number variation, human health, and disease. Lancet 374:340-350.

Wakeland, E.K., Liu, K., Graham, R.R., and Behrens, T.W. (2001). Delineating the genetic basis of systemic lupus erythematosus. Immunity 15:397-408.

Walport, M.J., Davies, K.A., and Botto, M. (1998). C1q and systemic lupus erythematosus. Immunobiology 199:265-285.

Walsh, F.S., and Doherty, P. (1997). Neural cell adhesion molecules of the immunoglobulin superfamily: role in axon growth and guidance. Annual review of cell and developmental biology 13:425-456.

van den Oord, E.J., Kuo, P.H., Hartmann, A.M., Webb, B.T., *et al.* (2008). Genomewide association analysis followed by a replication study implicates a novel candidate gene for neuroticism. Archives of general psychiatry 65:1062-1071.

Wandstrat, A., and Wakeland, E. (2001). The genetics of complex autoimmune diseases: non-MHC susceptibility genes. Nat Immunol 2:802-809.

Wang, W.Y., Cordell, H.J., and Todd, J.A. (2003). Association mapping of complex diseases in linked regions: estimation of genetic effects and feasibility of testing rare variants. Genetic epidemiology 24:36-43.

VanGuilder, H.D., Vrana, K.E., and Freeman, W.M. (2008). Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques 44:619-626.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., *et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature 420:520-562.

Webb, R., Wren, J.D., Jeffries, M., Kelly, J.A., *et al.* (2009). Variants within MECP2, a key transcription regulator, are associated with increased susceptibility to lupus and differential gene expression in patients with systemic lupus erythematosus. Arthritis Rheum 60:1076-1084.

Viallard, J.F., Pellegrin, J.L., Ranchin, V., Schaeverbeke, T., *et al.* (1999). Th1 (IL-2, interferon-gamma (IFN-gamma)) and Th2 (IL-10, IL-4) cytokine production by peripheral blood mononuclear cells (PBMC) from patients with systemic lupus erythematosus (SLE). Clinical and experimental immunology 115:189-195.

Widiger, T.A., and Trull, T.J. (1992). Personality and psychopathology: an application of the five-factor model. Journal of personality 60:363-393.

Voight, B.F., and Pritchard, J.K. (2005). Confounding from cryptic relatedness in case-control association studies. PLoS genetics 1:e32.

Wong, C.K., Li, E.K., Ho, C.Y., and Lam, C.W. (2000). Elevation of plasma interleukin-18 concentration is correlated with disease activity in systemic lupus erythematosus. Rheumatology (Oxford) 39:1078-1081.

Wong, C.K., Lit, L.C., Tam, L.S., Li, E.K., *et al.* (2008). Hyperproduction of IL-23 and IL-17 in patients with systemic lupus erythematosus: implications for Th17-mediated inflammation in auto-immunity. Clinical immunology 127:385-393.

Xing, C., Sestak, A.L., Kelly, J.A., Nguyen, K.L., *et al.* (2007). Localization and replication of the systemic lupus erythematosus linkage signal at 4p16: interaction with 2p11, 12q24 and 19q13 in European Americans. Hum Genet 120:623-631.

Yang, Y., Chung, E.K., Wu, Y.L., Savelli, S.L., *et al.* (2007). Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. Am J Hum Genet 80:1037-1054.

Yousefi, S., and Simon, H.U. (2007). Apoptosis regulation by autophagy gene 5. Critical reviews in oncology/hematology *63*, 241-244.

Zhernakova, A., van Diemen, C.C., and Wijmenga, C. (2009). Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nature reviews 10:43-55.