THE EXTENDED MHC HAPLOTYPES AND THEIR ROLE IN SARCOIDOSIS

Annika Wennerström

Helsinki University Biomedical Dissertations No. 191

Department of Medicine, Transplantation Laboratory,
Haartman Institute, University of Helsinki, Helsinki Finland
Helsinki Biomedical Graduate School,
League of European Research Universities

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine of University of Helsinki, for public examination in lecture hall 2, Haartman Institute,
On 31 January 2014, at 12 noon.

Finland 2014

Supervised by
Docent Marja-Liisa Lokki, PhD
Transplantation Laboratory,
University of Helsinki,
Helsinki, Finland

Reviewers
Professor Hannes Lohi, PhD
Research Programs Unit, Molecular Neurology, Faculty of Medicine,
Department of Veterinary Biosciences, Faculty of Veterinary Medicine,
University of Helsinki and Folkhälsan Research Center, Helsinki, Finland

Professor Pentti Tienari, MD, PhD
Research Program for Molecular Neurology
Helsinki University Central Hospital
University of Helsinki, Helsinki, Finland

Opponent
Docent Janna Saarela, MD, PhD
Institute for Molecular Medicine Finland (FIMM),
University of Helsinki, Helsinki, Finland

ISSN 1457-8433
ISBN 978-952-10-9724-9 (pbk.)
ISBN 978-952-10-9725-6 (PDF)

LASERPAINO
Helsinki 2014

ABSTRACT

A large part of the genomic arrangement regulating the immune system is located in the Major Histocompatibility Complex (MHC) region on chromosome 6p21.31. Human leucocyte antigen (HLA) genes, which encode the antigen presenting molecules, are located in MHC class I and II, while the MHC class III region contains non-HLA genes, e.g. butyrophilin-like protein 2 (BTNL2), tumor necrosis factor (TNF) and complement C4. Most genes in the MHC region play an important role in susceptibility to autoimmune and infectious diseases.

Sarcoidosis is a complex granulomatous disorder with varying a clinical presentation. A multitude of studies have reported strong associations between genes in the MHC and sarcoidosis. The structure of the MHC is characterized by a high degree of polymorphism, due to population-specific and highly conserved extended haplotypes, leading to a complicated pattern of linkage disequilibrium (LD). This makes dissecting disease associations within individual MHC genes difficult.

The primary aims of this Study were to study the overall polymorphism of the MHC region in Finnish subjects, to investigate MHC as a susceptibility locus for sarcoidosis and to find predictive markers for the prognosis of the disease.

In Study I, extended MHC haplotypes covering genes within MHC class I (HLA-A, -B), MHC class II (HLA-DRB1, -DQB1 and –DPB1) and MHC class III (e.g. TNF and BTNL2) were constructed and LD between the genes studied in a Finnish population sample. We discovered that the extended MHC haplotypes could be grouped, based on a similar functional variant (e.g. truncating protein) or a similar structure of the peptide binding site. In Study II, we replicated a previously reported HLA-DRB1*03:01 association with a favourable prognosis of sarcoidosis in Finland (P=0.007, OR=2.7). Using the extended MHC haplotype approach, we detected novel and independent variants associated with sarcoidosis (HLA-DPB1*04:02; P=0.003, OR=0.48) or the disease course of sarcoidosis (HLA-DRB1*04:01-DPB1*04:01; P=0.02, OR=3.07). In Study IV, we explored four genes in the MHC class III region (LTA, TNF, AGER, BTNL2) and HLA-DRA, and performed a replication and meta-analysis in four European populations. After adjusting for sex, population stratification and HLA-DRB1 alleles, four SNPs in the HLA-DRA/BTNL2 region were associated with non-Löfgren sarcoidosis; the strongest signal association was detected with rs3177928 (1.79x10-7, OR=1.90). In this Study, we also addressed guidelines for disease association studies dealing with immunogenetic data.

The results of the extended MHC haplotype analysis discovered a unique haplotype admixture in the Finnish population, which may have implications to MHC-related disease associations. Our sarcoidosis study provided new

insights to predicting the disease prognosis, with different distributions of MHC variants associating different patterns of progression, further promoting the importance of MHC region in sarcoidosis predisposition. Especially, the region covering the genes BTNL2 and HLA-DRA warrants further studies in larger samples and in different ethnic groups. To conclude, this Study showed the importance of studying extended MHC haplotypes in well-characterized samples, in order to understand the complex MHC structure and to distinguish variants' associations with the trait.

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:

I Wennerström A*, Vlachopoulou E, Lahtela L, Paakkanen R, Eronen K.T, Seppänen M, Lokki M-L (2013) Diversity of extended HLA-DRB1 haplotypes in the Finnish population, PLOS One, PLoS One. 2013 Nov 21;8(11):e79690. doi: 10.1371/journal.pone.

II Wennerström A*, Pietinalho A, Vauhkonen H, Lahtela L, Palikhe A, Hedman J, Purokivi M, Varkki E, Seppänen M, Lokki M-L, Selroos O and the Finnish Sarcoidosis Study Group (2012) HLA-DRB1 allele frequencies and C4 copy number variation in Finnish sarcoidosis patients and associations with disease prognosis, Human Immunology 73(1):93-100.

III Wennerström A*, Pietinalho A, Lasota J, Salli K, Surakka I, Seppänen M, Selroos Oand Lokki M-L (2013) Major histocompatibility complex class II and BTNL2 associations in sarcoidosis, Eur Respir J 2013; 42: 550–553.

IV Wennerström A*, Lahtela E, Anttila V, Grunewald J, van Moorsel C.H.M., Petrek M, Eklund A, Grutters J.C, Kolek V, Mrazek F, Padyukov L, Pietinalho A, Ronninger M, Seppänen M, Selroos O, Lokki M-L (2013) SNP variants in MHC are associated with sarcoidosis susceptibility and subphenotypes – a joint case-control association study in four European populations (submitted)

*First and corresponding author
The publications are referred to in the text by their Roman numerals.
These articles are reproduced with the kind permission of their copyright holders.
Some unpublished data is presented.

ABBREVIATIONS

aa	amino acid
AGER (RAGE)	receptor for advanced glycation endproducts
AH	ancestral haplotype
Ala	alanine
ANXA11	annexin A11
APC	antigen presenting cells
AS	ankylosing spondylitis
Asp	aspartic acid
BHL	bilateral hilar lymphadenopathy
BCR	B cell receptor
BMDW	Bone Marrow Donor Worldwide
BMT	bone marrow transplantation
bp	base pair
BTNL2	butyrophilin-like protein 2
CBD	chronic beryllium disease
CD	celiac disease
CEU	Utah residents with ancestry from northern and western Europe (a HapMap reference population)
CHR	chromosome
CI	confidence interval
CNV	copy number variation
CTins	CT insertion mutation
CWD	common, well documented
C4	complement component 4
C4A	acidic isotype of complement component 4
C4B	basic isotype of complement component 4
DNA	deoxyribonucleic acid
DZ	dizygotic
ECG	electrocardiography
EFI	European Federation for Immunogenetics
ENCODE	the Encyclopedia of DNA Elements
ERAP1	endoplasmic reticulum aminopeptidas
eQTL	expression quantitative trait locus
FIMM	Institute for Molecular Medicine Finland
GD	Graves' disease
GWAS	genome-wide association study
HARPS	Ambiguity Resolving primers
HLA	human leucocyte antigen
HSCT	hematopoietic stem cell transplantation
HWE	Hardy Weinberg equilibrium
HVR	hypervariable regions

h2	heritability (narrow-sense)
LD	linkage disequilibrium
LS	Löfgren syndrome
LTA	lymphotoxin alfa
MAF	minor allele frequency
mRNA	messenger ribonucleic acid
MHC	major histocompatibility complex
miRNA	microRNA
MS	multiple sclerosis
MZ	monozygotic
NGS	next-generation sequencing
NL	non-Löfgren syndrome
NLR	non-Löfgren syndrome, resolved sarcoidosis
NLP	non-Löfgren syndrome, persistent sarcoidosis
OR	odds ratio
RA	rheumatoid arthritis
RT-PCR	realtime-PCR
PBS	peptide-binding site
Pc	corrected P value
PCR	polymerase chain reaction
PCS	protein coding site
P9	peptide-binding pocket nine
S-ACE	angiotensin-converting enzyme
SBT	Sequence-Based Typing
Ser	serine
SSO	Sequence-Specific Oligohybridization
SSP	Sequence-Specific primer
QC	quality control
qPCR	quantitative polymerase chain reaction (PCR)
SLE	systemic lupus erythematosus
SNP	single nucleotide polymorphism
tag-SNP	tagging-SNP
TCR	T cell receptor
TF	transcription factor
TID	type I diabetes
TNF	tumor necrosis factor
TFIID	transcription factor II D
$\chi 2$	Chi-square

1 INTRODUCTION

The most polymorphic genetic region in the human genome, the major histocompatibility complex (MHC), has been associated with several inflammatory (e.g. sarcoidosis) and autoimmune diseases [e.g. type I diabetes (T1D) and multiple sclerosis (MS)], as well as graft failure [1]. The MHC region has a complex allelic structure with extended linkage disequilibrium (LD) and polymorphism, reviewed in [2]. The genes that encode the cell-surface antigen-presenting proteins, the human leucocyte antigen (HLA) genes, are located at the MHC region among many other genes related to immune response. The main role of the immune system is to recognize and distinguish self from non-self [3]. The previous studies on MHC region have increased our knowledge of the genetic risk factors for autoimmune and infectious diseases, shed light understanding the disease pathogenesis, identification of individuals with high risk, disease prognosis and suggested therapeutic approaches.

One of the MHC-related diseases is sarcoidosis. Sarcoidosis is a systemic disease of unknown etiology and with varying clinical course [4]. Based on a nation-wide survey performed in 1984, the crude prevalence of sarcoidosis in Finland was $30/10^5$ and the annual incidence $11/10^5$ [5,6]. Sarcoidosis is a complex disease, with interactive effects from the environment and multiple genes. Both linkage and association studies, including the recent genome-wide association studies (GWASs), have shown a strong role for MHC in susceptibility to sarcoidosis [7] with distinct disease phenotypes and populations. To explore the prognosis of sarcoidosis, the immunologically important MHC region offers opportunities to discover disease phenotype specific gene variants.

This study aimed to characterize the general polymorphism of the MHC region in Finnish subjects and in sarcoidosis patients to discover novel risk factors and to improve the prediction of disease course. Overall, we set out to understand the complexity of the MHC data to suggest guidelines on how the MHC disease association studies should be conducted, taking into account the special challenges in immunogenetic data analysis: the highly polymorphic nature of the MHC region, population stratification, and the resulting statistical issues.

2 REVIEW OF THE LITERATURE

2.1 THE HUMAN GENOME

The international scientific collaboration project, the Human Genome Project, launched the first draft of the sequenced human genome in the beginning of the 21st century [8,9]. The project estimated the number of protein-encoding genes to be approximately 21 000, coded by sequence that represents only a small fraction of the human genome (1.1-3 %) [10]. The majority of the human genome consists of functional non-coding regions, such as micro-RNA, repetitive sequences (e.g., tandem repeats associated with various diseases [11]), gene expression regulators, and regions of unknown function. The molecular basis of inheritance, the deoxyribonucleic acid (DNA), is located in the chromosomes at nucleus of every cell in the body (except mature red blood cells). Typically, the human genome consist of 22 pairs of autosomal chromosomes and two sex chromosomes (XX or XY), with one half of the pair of sister chromosomes inherited from the mother, the other one from the father.

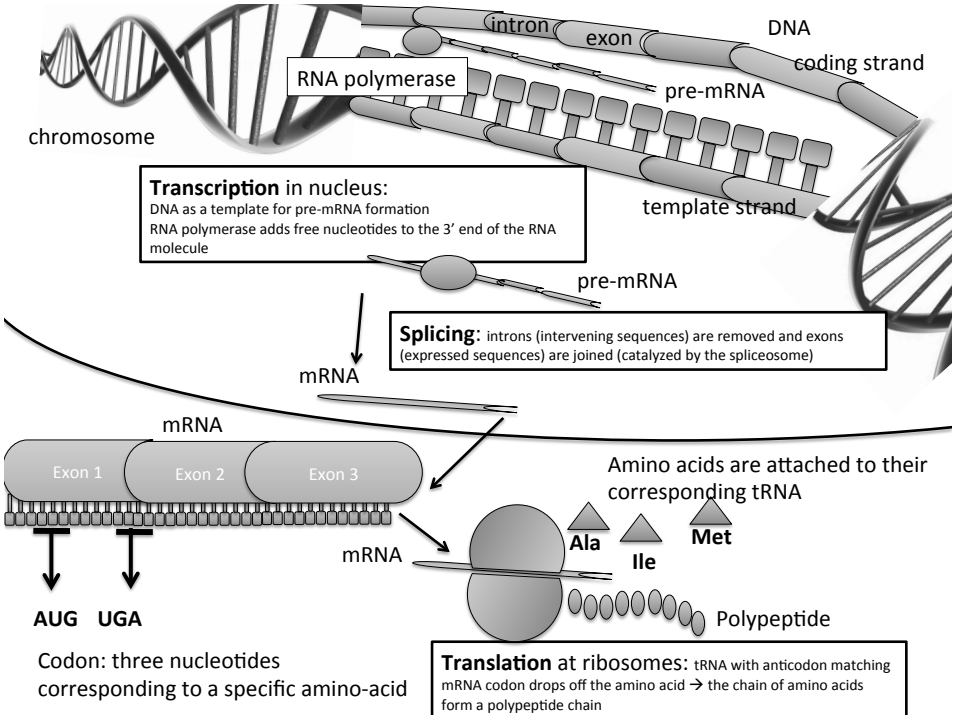


Figure 1 From chromosomes to protein. Protein-coding genes contain segments called the promoter, a number of introns and exons and the 3' region. Before mRNA (the transcript of the gene sequence) is translated into a protein at the ribosomes, it is processed in the nucleus as introns are removed and exons spliced together. tRNA, transfer RNA; mRNA, messenger RNA

The structure of genes can be roughly divided into four segments; promoter, exons, introns and downstream regulatory regions (3' end). The promoter region in the 5' prime side of the gene acts as a gene regulator binding transcription factor II D (TFIID) [12,13], and consists of a conserved TATA-box and a variable CpG-rich site [14]. Figure 1 shows the key factors in transcription and translation. Transcription of DNA to RNA occurs from 5' to 3' prime end. Exons comprise the gene code that will be transcripted into pre-messenger ribonucleic acid (pre-mRNA). The introns, which are not involved in the protein formation, are spliced off from the pre-mRNA to form mRNA. Translation takes place at the ribosomes, where transfer RNA (tRNA) transfers amino acids to form the polypeptide chain based on the mRNA code (corresponding with the DNA coding strand) (Fig. 1) [8,9].

2.1.1 GENOME REGULATION

The regulation of gene expression is a complex and strictly regulated process containing various factors and levels of regulation. Furthermore, the protein coding genes are differentially expressed and regulated in different cells. Any stage of gene regulation and expression (chromatin domains, transcription, RNA transport from nucleus to cytoplasm, translation, mRNA degradation) can be modified. Particular genes, involved e.g., in DNA replication and repair, are continuously expressed. However, the most genes are regulated occasionally and "turned off" by histone proteins that are tightly bound to the DNA [15]. In order for transcription to take place, the DNA needs to be unwound, e.g., by histone modifications. In addition, modifier genes that affect the expression of some alleles, external environmental factors, transcription factors (TFs) and microRNA (miRNA) regulate gene expression [15,16]. In addition to the promoter region, regulatory features exist, including the 5' untranslated region that controls the initiation of translation and enhancers that can be physically situated far away from the gene, [17]. Polymorphisms in the protein coding site (PCS) or regulatory elements (e.g., transcription factor binding sites) can alter the structure of the gene product or its function [18,19]. The gene regulation can be studied e.g., by exploring SNPs that associate with gene expression [expression quantitative trait locus (eQTL)] [20,21].

In recent years, the Encyclopedia of DNA Elements (ENCODE) project has discovered novel features of the human genome (e.g., novel non-coding RNA, novel transcription promoters and novel gene isoforms) that provide new insight into the concepts of gene and gene regulation [15,22-24]. Furthermore, the ENCODE project has suggested that at least 80% of the genome has some biochemical function (e.g., transcription or DNA regulation) in at least one cell type [22]. Interestingly, the ENCODE project has also estimated that over 60% of the genome is transcribed into RNA [15], although only a small fraction of the genome encodes proteins. The ENCODE project shows that the previously believed junk DNA is actually functional: The most of the non-

coding DNA is involved in gene's regulation suggesting a far more complex regulation process than previously thought [25].

2.1.2 GENETIC VARIATION

International consortia, including the 1000 Genomes Project [21,26], the International HapMap Project [27] and the ENCODE consortium [25] have explored the human variation by studying the rare variants, ancestral haplotypes and functional regions of the human genome, respectively, to understand the variance and function of the genome and its relation to the traits. In general, genetic variation involves changes to the base sequence of DNA. The average rate for *de novo* variation per generation is 1.20×10^{-8} per nucleotide which corresponds to approximately 70 novel single nucleotide variations per genome per generation [28]. It has been speculated that most of the polymorphism in the genome is rare, with a minor allele frequency (MAF) <1%. A polymorphism may involve a single base change, called a point mutation, or larger sections of DNA through deletions, insertions (Fig. 2), or translocations [25].

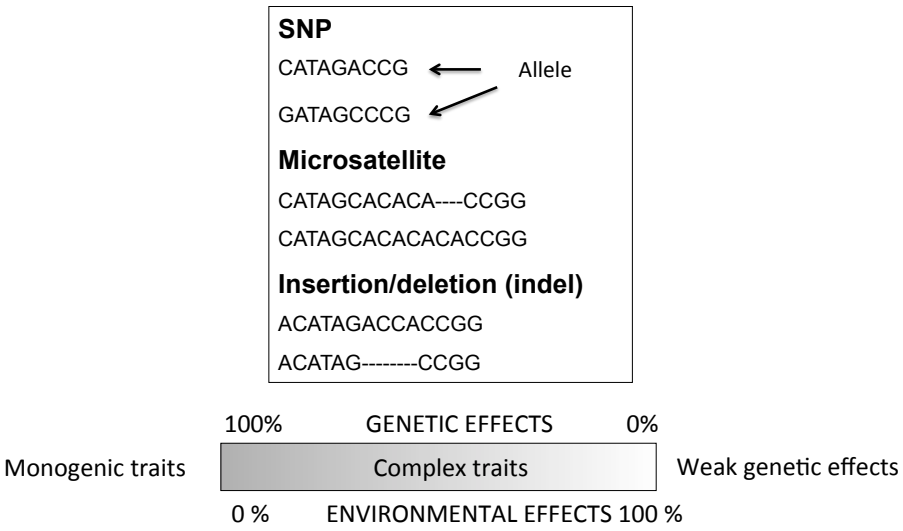


Figure 2 Monogenic traits are the result of variation in a single gene. Complex traits are the result of interplay between numerous genetic susceptibility factors and environmental factors. SNPs are the most common type of variation. The structural variations (e.g., microsatellites or insertion-deletion polymorphism) affect the length of the sequence. The alleles are the variants of a DNA sequence at locus (physical location of the variant), which can be bi-allelic (SNP) or multiallelic (HLA genes, copy number variations).

The most typical example of sequence variation is the single nucleotide polymorphisms (SNP) that are common in the genome, and are typically biallelic. Copy number variation (e.g. microsatellites) and insertion-deletion polymorphisms represent structural variation, which occur at much lower frequencies than SNPs [29]. A frameshift mutation is caused by the addition

or the loss of one or more nucleotides. A silent mutation does not have any effect on the resulting protein. In case of splice site mutation, an altered mRNA sequence is formed as the splicing of introns are damaged resulting in a non-functional protein [21].

Contradictory to monogenic traits, which are typically affected mainly by only a few genes, various traits are complex, with both genetic and environmental factors being needed for the occurrence of the trait (Fig. 2). The concepts of sequence variation, monogenic and complex disorders and allele are highlighted in Fig. 2. The effect of the alleles can be dominant (only one copy is required for a particular trait to be expressed), recessive (two copies of the allele are required for the trait to be expressed) or co-dominant (two dominant alleles both expressed). A genotype is referred to as the genetic makeup of a person, while a phenotype is the manifestation of the genotype (Fig. 2). When the same genotype leads to the same disease phenotype, in case of many monogenetic diseases, the penetrance is complete. However, if the penetrance is incomplete, all the allele carriers do not express the trait. Genetic heterogeneity means that different genotypes in different loci (in a same gene or different gene) result in the same outcome, phenotype. Contrary, different disease phenotypes can be the result of the same variation or different variation in the same gene [21]

Epigenetics can be defined by events that alter the phenotype but do not change the nucleotide sequence, such as posttranslational modifications of histones, the methylation of CpG islands, phosphorylation, and the expression of noncoding RNAs. Epigenetics can be considered to be a link between a genotype and a phenotype and contributes to the development of complex diseases. Furthermore, the mutation in nucleotide sequence may cause failure in the methylation sites and environmental triggers, such as carcinogens, may permanently alter epigenetic patterns [30-32].

Heritability is a mathematical estimate, which is defined as a degree of variation in a trait that can be explained by genetic factors [33]. Heritability is dependent on the population studied, typically measured as the narrow-sense heritability (h^2) and expressed as percentage value. Twins are typically studied for heritability estimates: On average the MZ twins share the same genome and the DZ twins share genes like normal siblings share (50% of the genome). If a disease is more correlated in MZ than in DZ twins, the genetic factors are the likely to influence the aetiology of the disease [33,34]. For example, the heritability (h^2) of the height in the Western countries is roughly 80%, indicating that 20% of the height variance can be explained by the environmental factors between the individuals (reviewed in [35]).

2.2 THE IMMUNE SYSTEM

The role of the immune system is to protect the host organism from potentially pathogenic agents or substances, including bacteria, viruses, toxins, or cancer cells, by recognizing harmful non-self-antigens and altered

self-antigens immediately with the innate immune system and with antigen specific adaptive immune response (Fig. 3). The structure and development of the immune system is unique for every individual. Numerous cell populations, mostly leucocytes, react against potential antigens through innate or by adaptive immunity mechanisms. Most immune system cells are T and B lymphocytes (i.e., T and B cells) or macrophages and neutrophils (reviewed in [36,37].

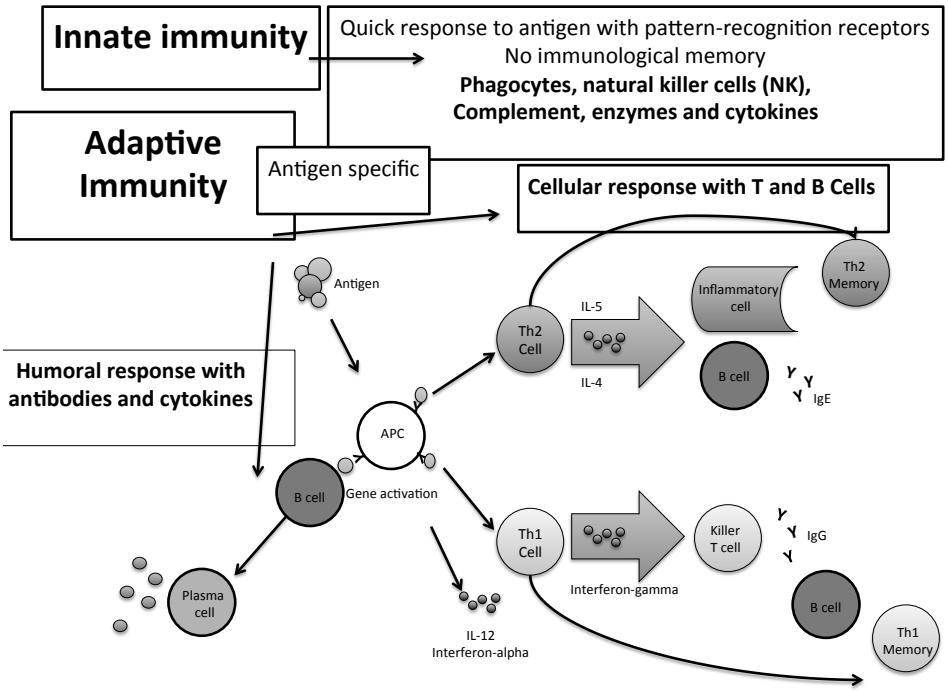


Figure 3 The immune system, typically divided into two categories, the innate and the adaptive immunity, operates through complex networks and pathways to defend the body against attacks by foreign invaders. Innate and adaptive systems work together to eliminate pathogens. The innate immunity system is quick in response to nonspecific microbe-specific molecules that are recognized by a given pattern-recognition receptor (PRR; e.g., toll-like receptors). The innate immunity uses phagocytosic cells, which release inflammatory mediators, natural killer (NK) cells and molecular components such as those involved in the complement cascade. The innate responses do not improve on repeated exposure to the antigen, as the adaptive responses do. The adaptive immune system is antigen-specific, where the antigen-presenting cell (APC) presents the particular antigen to lymphocytes. The adaptive immune systems have two adaptive mechanisms: cell-mediated immunity with T and B cells and humoral immunity with antibodies produced by plasma cells and cytokines. [36]

The innate system is among the first cells to encounter pathogens inside the physical protective barriers, skin and mucous membranes. The innate system recognizes non-self-structures typical to many pathogens with non-antigen specific pattern-recognition receptors (PRR), e.g., toll-like receptors [38]. In the adaptive immune system, the T cells are the key element in recognizing non-self via the T cell receptor (TCR) MHC –complex that is the

main effector of humoral responses. TCRs are transmembrane heterodimer molecules [39]. Antigen presenting cells (APC, e.g., dendritic cells and macrophages) bind intracellulary processed antigens and present antigen in the MHC molecule that is on the surface of the cell. The first contact with the pathogenic agents differentiates the naïve CD4+ T lymphocytes (Th0) into various populations, such as T helper (CD4+ Th1, Th2, Th17) or regulatory T cells (CD4+ Tregs), later regulated by epigenetic mechanisms via cytokine production and presence of certain cell types. The CD8+ T lymphocytes are cytotoxic T cells [40]. When the B cells recognize an antigen with the B cell receptor (BCR), B cells differentiate into plasma cells that secrete antibodies, which in turn destroy pathogens by activate complement cascades and recruiting other cells. Furthermore, each B lymphocyte clone can produce particular circulating antibodies (immunoglobulins IgA, IgE, IgG or IgM). The immunoglobulins are glycoproteins and all the classes have different functions (reviewed in [36]**.**

Autoimmune disease may develop as a result of altered balance between T regulatory lymphocytes and self-reactive T lymphocytes. The causality in autoimmune, and as well as inflammation traits, is the incorrect action of the immune system against self-tissue that should be normally tolerated. Typically, the negative selection prevents autoimmunity, removing autoreactive B lymphocytes in bone marrow and autoreactive T lymphocytes within the thymus. In addition, peripheral tolerance removes autoreactive lymphocytes based on incomplete activation signals. The autoimmune diseases can be classified by the mechanism of tissue damage. In systematic autoimmune diseases, such as systemic lupus erythematosus (SLE) [41], various tissues are affected and the autoimmunity is mediated both by T and B lymphocytes [42]. In T-cell-mediated diseases, multiple sclerosis (MS) or type I diabetes (T1D), a particular tissue is attacked by the autoreactive cells [43-46]. Specifically, in the case of T1D, the immune system does not tolerate self-pancreatic β-cells in the Islets of Langerhans resulting in loss of insulin secretion [47,48].

2.3 THE MAJOR HISTOCOMPATIBILITY COMPLEX (MHC)

The human MHC region is located on the chromosomal region 6p21.31, spanning approximately 4 megabases of DNA. The first human MHC sequence was released in 1999 [49]. The MHC contains over 200 genes, which makes it one of the most gene-rich regions in the genome. In addition, compared with other genomic regions, the MHC shows the most variability characterized by high linkage disequilibrium (LD), and is rich in recombination hotspots and population specific features [50]. The variability of the MHC region is suggested to be strongly influenced by selection, e.g., pathogen-driven balancing selection [51,52].

Genetic structure of the human MHC region can be roughly divided into three sections: the MHC class I, II and III (Fig. 4). It has been speculated that the region was developed from a block of duplicated immune systems [50,53]. The traditional Human leucocyte antigen (HLA) genes, *HLA-A*, *-B* and *–C*, are located in the MHC class I (i.e., HLA class I) and *HLA-DRB1*, *-DQB1* or *-DPB1* are located in MHC class II (i.e. HLA class II) [54].The structure of MHC class I and II molecules is presented in Fig 4.

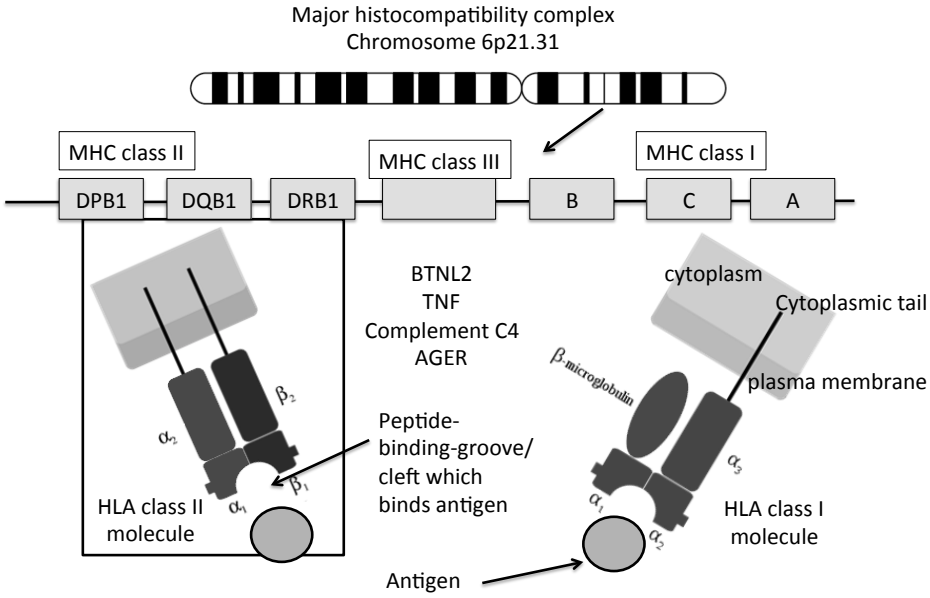


Figure 4 The human MHC region in chromosome 6 encodes antigen presenting HLA molecules (in MHC class I and II) and critical mediators of innate immunity and many non-immune genes in MHC class III genes e.g., *BTNL2* (butyrophilin-like 2), *TNF* (tumor necrosis factor), *complement C4*, *AGER* (advanced glycosylation end product receptor). The HLA class I molecules consist of an α-subunit that includes the polymorphic peptide-binding cleft (formed by α_1 and α_2) and of the non-polymorphic β-subunit (β_2) which is encoded by the *beta-2 microglobulin* (*B2M)* gene on chromosome 15[58]. The MHC I α3 domain spans though plasma membrane and interacts with T cell receptor (TCR). The HLA class II molecules are heterodimers formed by the α-subunit (encoded by *HLA-DRA, -DQA1* or *-DPA1*) and by highly polymorphic β-subunit (encoded by *HLA-DRB1*, *-DQB1* or *–DPB1).* In HLA class II, the antigen binding cleft is encoded by α_1 and β_1 domains (reviewed in [59])**.**

A sequence of the newly defined xMHC was completed in 2003 as part of the sequencing of the entire human chromosome 6 and now covers a total of 7.6 Mb on the short arm of this chromosome. The extended MHC (xMHC; 7.6 Mb) was sequences in 2003 and includes the traditional HLA as well as the genes outside of MHC class I and class II (e.g. MHC class III region) [55]. Roughly, the xMHC can be divided in to the extended MHC class I (*HIST1H2AA* to *MOG*; 3.9 Mb), the classical MHC class I (*C6orf40* to *MICB*; 1.9 Mb), the classical MHC class III (*PPIP9* to *NOTCH4*; 0.7 Mb), the classical MHC class II (*C6orf10* to *HCG24*; 0.9 Mb) and the extended MHC class II (*COL11A2* to *RPL12P1*; 0.2 Mb) (reviewed in [50]). The MHC class III region

located between the MHC classes I and II is the most gene rich region of the MHC. There are three types of genes in the MHC class III region; immune response related genes (e.g., *TNF*), non-immunological genes, and genes that are not expressed (pseudogenes). The genes in the MHC class III are typically tightly linked with other MHC genes with potential functional interaction with the HLA genes [54,56,57].

The HLA genes are differentially expressed in different cell populations. The HLA class I molecules are expressed on most nucleated cell surfaces and present intracellular pathogen antigens to CD8+ cytotoxic T cells. HLA class I molecules bind short peptide fragments (the majority of these peptides are length of 9) generated in the cytosol by proteasomal degradation [60,61]. These short peptides anchor into the end of HLA class I peptide-binding groove[62]. If the complex activates, the presenting cell is destroyed by apoptosis.

Each HLA class I molecule is able to bind between 1000 and 10 000 peptides and each HLA class II molecule to bind approximately 2000 peptides [63]. The MHC class II molecules are presented on antigen presenting cells such as e.g., dendritic cells. The peptide binding-groove of the HLA class II molecules is open that enables variable size of peptides to bind into the molecules (typically size of the peptides are 8-15 amino acids) [64]. HLA class II molecules typically bind peptides from endocytosed proteins and present the peptides to CD4+ helper T cells. In addition, HLA class II molecules regulate self-tolerance immunity presenting purified peptides from class II molecules that are derived from cytosolic self-proteins [65]. The activation of the MHC class II CD4+ complex is presented in Fig. 3 [50].

2.3.1 MHC DIVERSITY BETWEEN SPECIES

In humans, the MHC region is often referred as the human leukocyte antigen (HLA) region and the antigen processing and presenting genes are called the HLA genes [49]. Primate species, especially higher primates (chimpanzees and gorillas) have similar MHC structures compared with humans. However, the largest differences between human and higher primates are the position of polymorphism, the unique HLA-A –related gene in chimpanzees and the distict variability in MIC-genes [66].

A detailed genetic structure of MHC region has been published several vertebrates such as, mouse (in chrom. 17; [67], rat (RT1 complex in Chr 20; [68], dog (DLA complex in Chr 12; [69]**)** and chicken [70]. The function of MHC and the architecture of the antigen-presenting molecules are conserved between species [71]. However, in other species the MHC region (especially MHC class I) differs extensively in both the number of genes and their arrangement [66]. For example, the MHC B-complex in chicken encodes only two genes each in class I and class II [72], and the mouse and rat MHCs [73,74]lack MIC-related genes altogether. In addition, MHC genes are regulated differentially between species; for example, the mice MHC class II molecules are not expressed on the cell surface of the T cells [75].

Translocation of the MHC in zebrafish shows the extent of the variation: MHC class I, II, and III are not linked as a complex as they are in many other vertebrates, and the structure of MHC class II varies extensively [76,77].

2.3.2 COMPLEX NOMENCLATURE OF THE HLA ALLELES

The HLA genes show extensive polymorphism and distinct variety among population. In contrast to the non-HLA genes, which have a modest number of alleles, a particular HLA gene can have up to 3000 alleles. There are some exceptions, e.g., *HLA-DRA,* which has only seven identified alleles. *HLA-B* and *HLA-DRB1* are the most polymorphic HLA class I and class II genes, respectively (IMGT/HLA database [78], Fig. 5). To compare, number of the canine MHC (DLA) alleles are 100 for *DLA-DRB1*, 26 for *DLA-DQA1* and 60 for *DLA-DQB1*[79].

The HLA alleles are co-dominantly inherited, i.e. both alleles are expressed unless a mutation occurs. The HLA null alleles are normally not expressed at the cell surface (as a serologically detectable product) due to variation in the sequence e.g., frameshift mutation leading into premature stop (*HLA-A*01:04N*) [80]. The number of alleles has increased dramatically in recent years (Fig. 5) [78]. However, only 10 % of these alleles have been fully sequenced and can be described as common, well documented (CWD) HLA alleles [81].

Gene	Alleles	Proteins	Null alleles
HLA class I			
HLA-A	2,432	1,740	117
HLA-B	3,086	2,329	101
HLA-C	2,035	1,445	57
HLA class II			
HLA-DRA	7	2	0
HLA-DRB1	1,375	1,020	27
HLA-DQA1	51	32	1
HLA-DQB1	459	303	13
HLA-DPA1	37	19	0
HLA-DPB1	193	160	6

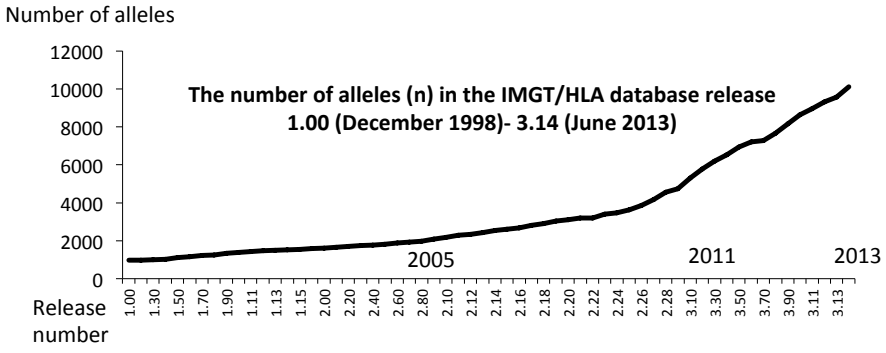


Figure 5 Statistics of HLA class I and class II alleles (Release 3.14/2013). The number of HLA alleles in the IMGT/HLA database release from 1998 to 2013. (http://www.ebi.ac.uk/imgt/hla/). Several alleles can encode the same protein. Null alleles are not expressed on the cell surface. [78]

The polymorphic HLA genes are classified according to the HLA nomenclature that is standardized and regularly updated by an international committee [IMGT/HLA Database (http://www.ebi.ac.uk/imgt/hla/)], and describes the gene-wide haplotypes of particular HLA gene [78,82,83]. Initially, the characterization of the HLA loci was needed for transplantation purposes. As presented in Fig. 6, the allele designations for each gene (e.g., *HLA-DRB1**) are a set of numbers separated by digits. The first field (first two digits) describes the allele family and the low-resolution allele, typically corresponding to serological antigens (e.g., *HLA-DRB1*01*). The following fields describe the amino-acid sequence of the gene product (*HLA-DRB1*01:01;* high-resolution allele) and any synonymous nucleotide substitutions in exons (*HLA-DRB1*01:01:01*). The last fields describe the polymorphism observed in the introns or in the 5′ and 3′ untranslated regions (*HLA-DRB1*01:01:01:01*). The suffix at the end of the digits describes the level of protein expression, ("N" i.e. null allele). HLA genotyping with "high resolution" (typically referred as typing with 4-digits, e.g. *HLA-DRB1*01:01* and resolving polymorphism in the peptide binding region) produces lower levels of ambiguity than typing with "low or medium resolution" (two-digit resolution, *HLA-DRB1*01*) [78,84].

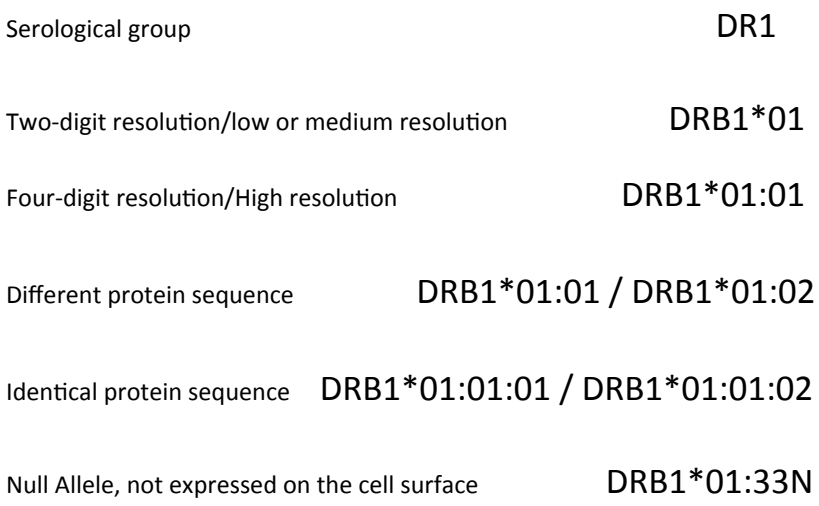


Figure 6 Complex HLA nomenclature [Adapted from IMGT/HLA Database (http://www.ebi.ac.uk/imgt/hla/)].

2.3.3 HLA TYPING METHODS

Traditionally HLA alleles were detected using serological techniques with the HLA-specific alloantisera and/or monoclonal antibodies. After the development of the DNA sequencing technology in the 1980s, HLA laboratories have mainly utilized DNA-based typing techniques, such as Sequence Specific Oligohybridization (SSO), Sequence Specific primer (SSP) typing and Sequence Based Typing (SBT). Typically, the polymorphisms in regions that encode the polymorphic peptide-binding site of the HLA

molecules are genotyped, specifically exon 2 for MHC class II loci, and exons 2 and 3 for MHC class I loci [84,85].

The genotyping of HLA genes and allele calling is challenging as the HLA alleles may differ only by a single base, and homological segments and pseudogenes complicate alignment with the reference sequence. The choice of HLA allele typing method depends on various aspects, such as how deep a resolution of the HLA allele is needed, how many samples are typed simultaneously and the cost of the method. With SBT and SSO, high throughput HLA typing can be performed. Luminex technology uses the SSO method where the DNA is hybridized by beads carrying an SSO probe. Fast but laborious, SSP typing uses many different allele or allele group-specific PCR primers in multiple PCR reactions. With SBT, the complete nucleotide sequence at the selected gene region is sequenced, offering a possibility to identify novel HLA alleles, which is not possible with SSO or SSP. However, the result of SBT, a hemizygous sequence, requires specialized allele assignment software for sequence alignment with the HLA allele reference library (IMGT database) [78].

In case of allele ambiguity (i.e., multiple allele combinations have the same sequence and the phase is not known), either heterozygous Ambiquity Resolving primers (HARPS) or sequencing of additional exons or introns are needed. By sequencing multiple exons of HLA genes, the potential amount of ambiguities increases. One approach for addressing allele ambiguity is to use the G groups (i.e., all class II alleles that share the same exon 2 nucleotide sequence and class I alleles that share the same exon 2 and 3 sequence) and the P groups (i.e., all alleles that encode the same peptide-binding region). The HLA allele ambiguities can also be resolved by using population-specific HLA allele distributions, known HLA haplotypes, and categorization of the HLA alleles into common and rare [78].

Today, newer HLA typing techniques are used in research, including high-throughput microarrays, HLA imputation and next-generation sequencing (NGS). It is challenging, even impossible, to design probes for HLA-alleles in high-throughput microarrays to cover both the ethnic polymorphism and within gene polymorphism of the HLA [86]. Thus, microarray-based SNP typing is not suitable for typing the HLA genes directly. However, SNPs [i.e. tagging-SNP (tag-SNP)] located away from the hotspots have been successfully used for imputing the traditional HLA alleles (or amino acid variants), if the SNP is in strong LD with a particular HLA allele [86]. Imputation is a statistical process where the SNP genotypes are used to infer missing HLA allele data. Furthermore, a tag- or proxy SNP represents the surrounding region, thus therefore only the SNP need to be genotyped. The imputation of HLA alleles has been shown to be a valuable tool for screening certain HLA alleles, e.g., *HLA-B*57:01* [87,88] and *HLA-DQB1* alleles for celiac disease [89] (Table 1), and boosted further development of imputation algorithms for GWASs (e.g. HLA*IMP [90,91], HIBAG [92], SNP2HLA [93]). The importance of independent validation of tag-SNPs in another population, and the use of population-specific reference panels for GWAS imputation is

important [82,89,94], given the population-specific LD patterns and polymorphisms in the MHC region,

Table 1. *Tag-SNPs for DQ-DR molecules associated with celiac disease [82,89,94]*

tag-SNP(s)	DQ		*DQ*-haplotype	*DRB1* allele
rs7454108	DQ8		DQA1*03:01-DQB1*03:02	DRB1*04
rs2395182, rs7775228, rs4713596	DQ2.2	DQ2 heterodimer in trans	DQA1*02:01-DQB1*02:02	DRB1*07:01
rs4639334	DQ7	DQ2 heterodimer in trans	DQA1*05:05-DQB1*03:01	DRB1*11/*12
rs2187668	DQ2.5	DQ2 heterodimer in cis	DQA1*05:01-DQB1*02:01	DRB1*03:01

NGS can be utilized for whole-genome or RNA-sequencing; however, the method is still too expensive for large-scale studies. NGS offers major advantages, including the absolute resolution of the alleles in long HLA haplotypes with phase information and the lack of ambiguous allele combinations [95]. However, the short reads in NGS (ranging from 25 to more than 500 bases) are typically insufficient for unambiguous assembly [96]. Indeed, large-scale reference panels for multiple global populations are warranted, as the de-novo assembly still needs improvement [2,97]. In addition, NGS has been hampered with structural variation (i.e., copy number variation and repetitive sequences) [98]. The four major sequencing platforms, Roche GS 454 FLX, Illumina MiSeq/HiSeq, PGM Ion Torrent and Pacific Biosciences, have different approaches depending on the instruments used, but with specialized analytical tools (e.g., Conexio Genomics (Fremantle, Western Australia, Australia) and GenDx (Utrecht, The Netherlands)), the genotyping results have been shown to be alike. Interestingly, a recent publication showed that the nanochannel genome mapping approach (with Illumina's HiSeq 2000) enabled to create MHC map that was consistent with the IMGT/HLA reference map. However, as only two MHC haploid clone libraries (PGF and COX) were used, more studies are needed, before using the approach in HLA typing [99]. In the near future, the NGS typing of HLA genes may become a cost-effective method in the field of histocompatibility [100-102].

2.3.4 THE CONCEPT OF THE HLA ALLELE AND MHC HAPLOTYPE

The HLA alleles differ in approximately 10-26 nucleotides from each other depending on the locus [103]. The peptide binding preferences for each HLA allele associate directly with the arrangement of the amino acid sequence that forms the peptide-binding site. Most of the polymorphism is observed in the region encoding the peptide-binding site (PBS), as shown in Fig. 7 (a part of *HLA-DRB1* exon 2) [104].

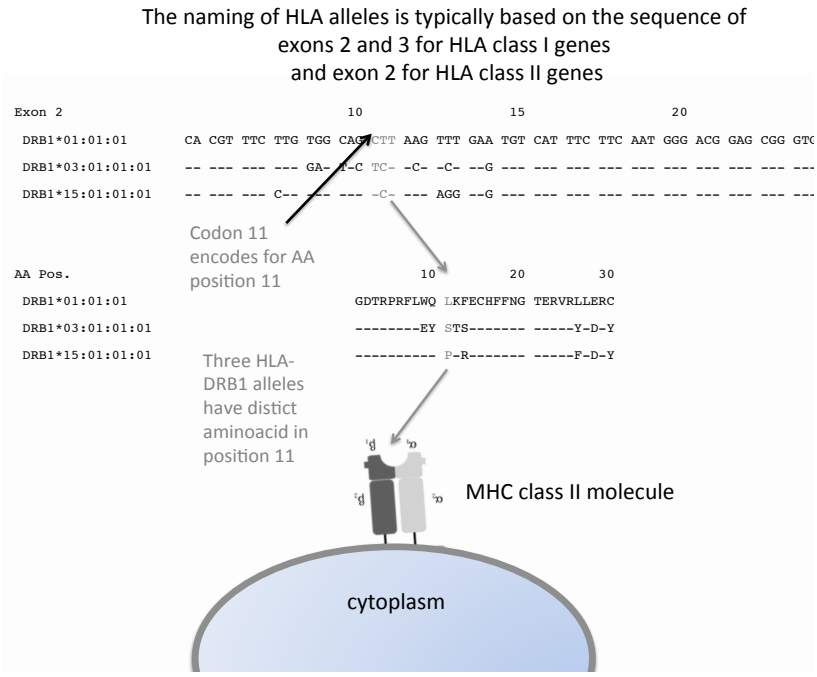


Figure 7 The allele calling of HLA class II is typically analysed based on the nucleotide sequence of exon 2. A part of the coding sequence of *HLA-DRB*1 is presented here. Codons 9-14 are highly polymorphic and encode the peptide-binding site (part of pockets 4, 6 and 9). Identities with the consensus (*HLA-DRB1*01:01:01)* are indicated by dash. *AA=amino acid*. [Adapted from [78]IMGT/HLA Database (http://www.ebi.ac.uk/imgt/hla/)] [78].

Most of the HLA alleles are rare and have been reported less than three times. The rest of the HLA alleles have low-to-medium frequencies, with the exception of *HLA-DPB1*, where only a few alleles account for the majority of the genes. Typically, the majority of the different alleles of *HLA-A, -B, -C, -DRB1*, and *-DQB1* are evenly distributed between different populations. For example, in most populations, approximately 20 alleles within *HLA-DRB1* account for the > 90 % of all alleles found in the population [81,105]. Due to the LD block structure (where nearby genomic fragments are inherited together), some HLA alleles and haplotypes are found more frequently than expected in the population. In addition, it has also been shown that the same allele can be generated in two or more independent events (i.e., convergent evolution). [104]

The term of MHC haplotype is used to describe the cis phase of the genes' alleles on the same chromosome (Fig. 8 and 9). The haplotype structure between closely related MHC genes can be broken by gene conversion or recombination during meiosis. In recombination, chromosome homologs are wound tightly around one another, and genetic segments are exchanged (in a process called crossing over) [15]. Figure 9 shows how MHC genes are inherited as a haplotype block without any recombination as well as with the recombination.

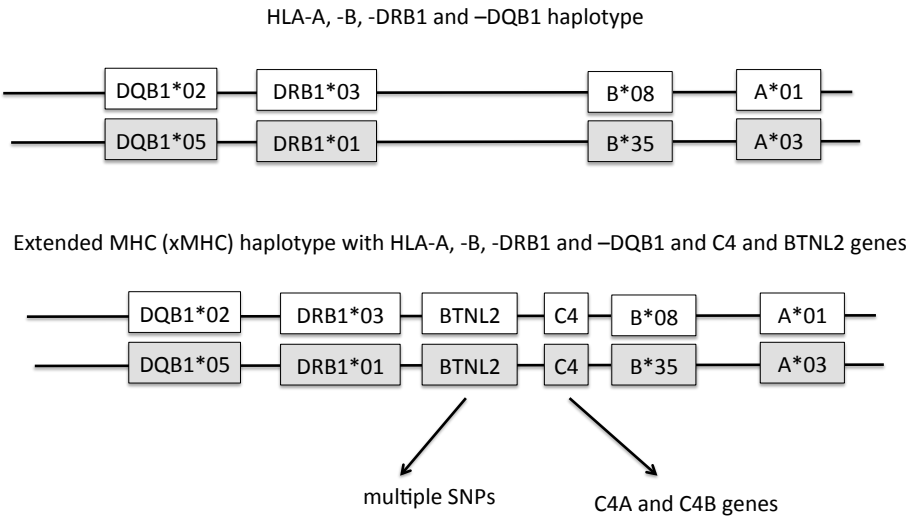


Figure 8 A *HLA-A, -B, -DRB1* and *–DQB1* haplotype is shown with the traditional HLA alleles (haplotypes *A*01-B*08-DRB1*03-DQB1*02* and *A*03-B*35-DRB1*01-DQB1*05*) and with using the extended MHC haplotype approach (haplotypes *A*01-B*08-C4-BTNL2-DRB1*03-DQB1*02* and *A*03-B*35-C4-BTNL2-DRB1*01-DQB1*05*).

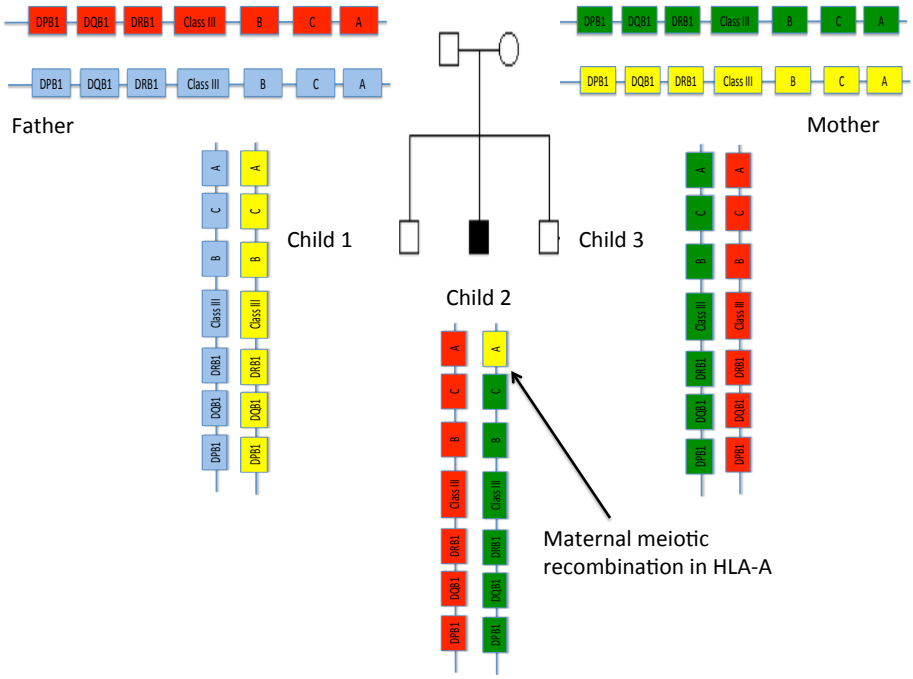


Figure 9 The inheritance of MHC haplotypes with (child 2) and without (Child 1 and 3) maternal recombination.

Most commonly, the phase of the HLA genes (haplotype) is estimated probabilistically using LD. When alleles of different contiguous loci occur together more often than expected from a random distribution according to their individual frequencies, the neighbouring loci are said to be in LD. The LD blocks are characterized as genetic segments with low recombination

[106]. The strength of LD is traditionally expressed by r^2 or D'. For multiallelic genes, one might argue that there are no good measures available, however, it has been suggested that, D' estimates the strength of LD (>0.80, strong LD; 0.8-0.5, moderate LD; 0.5-0, weak LD) better than r^2 [107]. If a meiotic recombination occurs (Fig.9), the new haplotype is formed from a combination of the parent's heterozygous haplotypes. Recombination hotspots have been described in the MHC region, denoting locations in the genome where the possibility of a meiotic recombination occurring is higher than the average. Known recombination hotspots occur in the MHC class I between *HLA-A* and *–C* and between *HLA-B* and MHC class II, and within class II (near *HLA-DPB1*). It is unlikely that recombination will occur between *HLA-DRB1* and *–DQB1*. The location of recombination hotspots and the length of LD blocks are population-specific [50,108].

2.3.5 HLA POLYMORPHISM IN POPULATIONS

The MHC allele and gene frequencies show strong differentiation between populations. Individuals who are heterozygous in their MHC alleles have a wider peptide-binding repertoire and therefore have a higher likelihood to respond to a wider range of pathogens [109]. The extensive population variability of the MHC genes may be a consequence of their functional ability to respond to more pathogen variants (i.e., pathogen-driven selection), as the HLA molecules play a critical role in immunity. However, pathogen-driven selection is not well understood. In addition, selective pressures (negative frequency-dependent selection and fluctuating selection) or a recent admixture of populations may offer explanation for the diversity of HLA alleles. A recent study confirmed that the geographical distances can be used to predict the HLA genetic diversity [109]. It has also been suggested that HLA class I and class II molecules have district evolution mechanisms [110,111]. The analysis of the distribution of HLA alleles has shown evidence of convergent evolution events among the HLA alleles, suggesting that some alleles are found in distinctive haplotypes in different populations. Typically, HLA alleles are related by descent from a common ancestral sequence (e.g., *HLA-DRB1*15:01* and **15:02*), but in some cases the alleles have developed independently. As an example, is *HLA-DRB1*08:04* haplotypes have with district HLA-DQB1 and –DQA1 alleles and *B*52:01:01* and *B*52:01:02* alleles [109].

The polymorphism of the HLA alleles in different population has been used to study the genetic distances between populations [112]. The study of HLA diversity in outbred populations has illustrated, for example, only two ethnicity-specific haplotypes for Europeans (*HLA-DRB1*08:01:01* and **16:01:01*) and several population-specific alleles for Asians, Africans and Native Americans. In addition, some rare haplotypes are more common in inbred populations, such as the extended *A*26:01-C*12:03-B*38:01-DRB1*04:02-DQB1*03:02* haplotype, which has a frequency of over 10% in the Ashkenazi Jewish population [104]. Similarly, by studying different dog

breeds, a restricted diversity of DLA alleles (due to e.g., inbreeding) has been observed within each dog breed, as well as a similar diversity of haplotypes between different dog breeds [79,113].

The studies have shown that HLA frequency data is highly correlated with geographic distances, illustrating a north-to-southwest axis. Extreme distributions of HLA diversity have been reported in subpopulations, such as the Norwegian Sami. The Norwegian Sami are genetically more closely related to the Finnish population than to other Norwegians. In addition, it has been found that the Finns themselves have a distinctive population substructure compared with other Europeans. The study of the LD patterns between MHC alleles in different populations provides a view to evolutionary relations between alleles and illustrates shared MHC blocks between alleles [104,109].

Collaborative studies including the International Histocompatibility Workshops and HLA-NET have examined the distribution of HLA alleles in different worldwide populations. Many of these populations' HLA data are available in the AlleleFrequencies.net –database [114] (http://www.allelefrequencies.net) or the dbMHC database (http://www.ncbi.nlm.nih.gov/projects/mhc). In Fig. 10, the frequency of *HLA-DRB1*15* from the former is shown [114].

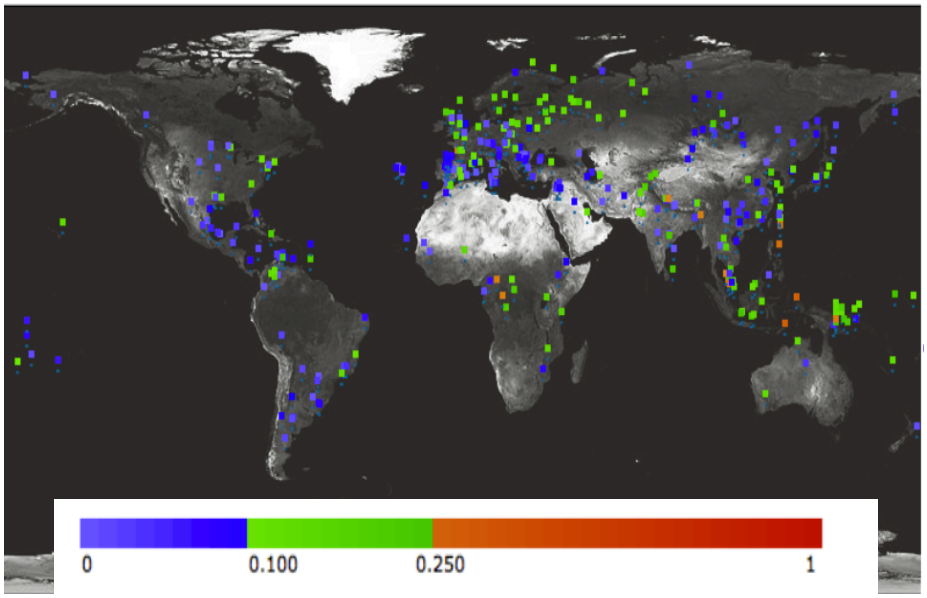


Figure 10 The *HLA-DRB1*15:01* allele frequency in the world based on the distribution of HLA alleles imported from the AlleleFrequencies.net –database (adapted from http://www.allelefrequencies.net)[114]

2.3.6 MHC AND TRANSPLANTATION

The HLA genes play an important role in stem cell transplantation and self vs. nonself –recognition in organ transplantation [115]. HLA-matching between the donor and recipient is critical in hematopoietic stem cell transplantation (HSCT) and bone marrow transplantation (BMT) for a successful outcome. The major complication in HSCT is an occurrence of graft-versus-host disease.

Solid organ transplants are not as sensitive for HLA mismatching as the HSCT, but graft rejection and loss are associated with HLA mismatching or mismatches.

The main objective is to find a donor with the same *HLA-A, -B, -C, -DRB1, -DQB1* alleles. A 10/10 allelic match represents that all the alleles are identical with donor and recipient [116,117]. However, the variability of the HLA complex restricts the number of possible donors and the most ideal donor, an HLA-identical sibling, is not always available. The likelihood of finding a suitable unrelated donor depends on the patient's racial and ethnic background [118]. Knowledge of the HLA diversity within each geographical region is used in finding a suitable donor. In national bone marrow (stem cell) registries and in the Bone Marrow Donor Worldwide (BMDW) databank (http://www.bmdw.org), millions of volunteer stem cell donors are registered. Due to the difficulties in finding an HLA-matching donor, many patients will be transplanted with a graft from an HLA-mismatched donor. Here, the challenge is to find the most optimal mismatched donor and to predict whether the recipient will being to make antibodies against the donor, as immunogenicity is not the same for every individual and certain mismatches are more permissible than others [119]. Previous studies have also suggested that MHC class III and *HLA-DPB1* regions are especially important in HSCT and influence on the outcome of the transplantation [120-122].

2.3.7 MHC DISEASE ASSOCIATION STUDIES

Previous GWASs have pointed out the importance of MHC in disease association studies (GWAS Integrator, http://genome.ucsc.edu, [123]). HLA has been confirmed as the major genetic factor for many autoimmune and inflammatory mediated diseases, including celiac disease (CD) [94,124-126], Graves' disease (GD) [127-129], T1D [130-134], MS [44,135], rheumatoid arthritis (RA**)** [136-140], ankylosing spondylitis (AS) [141]**,** systemic lupus erythematosus (SLE) [142,143] and narcolepsy [144] (Table 2). Autoimmune diseases, most often chronic and impossible to cure, affect up to 4% of the population in industrialized countries [56]. GWAS have revealed novel disease associations such as schizophrenia and Parkinson disease [145] and indicated additional susceptibility loci for autoimmune and inflammatory diseases outside MHC [50,86]. In addition, a growing number of the MHC associations with drug hypersensitivity reactions have been reported, including abacavir hypersensitivity syndrome (association with the abacavir and *HLA-B*57:01*) and the Steven-Johnson syndrome (association with carbamazepine and *HLA-B*15:02*) [146].

Several autoimmune, infectious and inflammatory diseases share some common predisposing MHC loci, as well as disease-specific markers [147]. The different associations with different disease outcomes may be due to different antigens and the presentation of different disease-specific autoantigens [56]. A frequently shared disease susceptibility allele, *HLA-DRB1*04* (or haplotypes including the allele) has been associated with several

diseases (e.g., RA, T1D [130,136]). T1D and celiac disease share genetic susceptibility markers suggesting common genetic background in HLA for these two autoimmune diseases [148,149]. Furthermore, many autoimmune diseases (e.g., GD, SLE) have been associated with *HLA-DRB1*03* or with the haplotype 8.1, typically referred as the autoimmunity haplotype [ancestral haplotype (AH) 8.1)], noticeably important in disease predisposition [150]. Table 2 describes the MHC associated diseases and conditions more thoroughly.

Table 2. *Disease associations of HLA alleles and their combinations*

HLA association	Disease	Reference
HLA-A*29	Birdshot retinopathy	[151]
HLA-B*27	Ankylosing spondylitis (AS)	[141,152,153]
HLA-C*06	Psoriasis vulgaris	[154]
HLA-DRB1*01	Coronary artery disease (CAD)	[155]
HLA-DRB1*03	Addison's disease (AS)	[156]
	Chronic active hepatitis	[157]
	Graves Disease (GD)	[152,158]
	Type I Diabetes (T1D)	[130]
	Myasthenia gravis	[159]
	Sjögren's syndrome	[160]
	Systemic lupus erythematosus (SLE)	[161]
	Sarcoidosis	[162]
HLA-DRB1*04 /	Rheumatoid arthritis (RA)	[136]
Shared epitope #	Rheumatoid arthritis (RA)	[137,138]
HLA-DRB1*04	T1D	[130]
HLA-DRB1*08	Systemic lupus erythematosus (SLE)	[143]
HLA-DRB1*15	Multiple sclerosis (MS)	[44,135]
	Narcolepsy	[144]
	Sarcoidosis	[163,164]
HLA-DQB1*02, *03:02	Celiac disease (CD)	[124,125,165]

*Shared epitope 70-74 QKRAA/QRRAA/RRRAA (*DRB1*01:01, *01:02, *04:01, *04:04, *04:05, *04:08, *10:01, *14:02*)

Furthermore, T1D has been shown to associate with peptide binding pocket nine (P9) of the DQB1 molecule, where β57 shows both susceptibility [alanine (Ala) or serine (Ser)] and protective associations [aspartic acid (Asp)] [46]. The β57 is a critical residue of the *HLA-DQB1* involved in antigen presentation and TCR interaction [46,133,166]. In addition, RA has shown evidence of association with the shared epitope within the *HLA-DRB1* peptide domain, at positions β70-β74 [138]. The association between antigen binding pockets' residues and disease sheds light on the importance of the particular antigen binding to pocket and genetic susceptibility.

The main trigger in autoimmune diseases is probably antigen presentation and T cell activation [167]. The immune response to a particular pathogen may depend on the MHC alleles carried by an individual. Many of the HLA-

alleles that are associated with a disease are common in healthy members of the population, suggesting that other genes and/or environmental factors are needed for disease manifestation. In addition, it has been suggested that the HLA risk variants might confer an evolutionary advantage [168].

The HLA associations with various diseases are typically much stronger than is typical for GWAS, showing high odd ratios (OR) for predisposing markers and low OR for protective markers. As an example, the HLA susceptibility haplotype for T1D has an OR of 3.64 (*DRB1*03:01-DQA1*05:01-DQB1*02:01*) and the protective haplotype has an OR of 0.03 (*DRB1*15:01-DQA1*01:02-DQB1*06:02*) [169]. However, while this makes them easy to detect in a GWAS, the mechanism behind HLA association is not well understood. It has been suggested that different dog breeds represent an excellent comparative model for human diseases. A good example is diabetes, which has a similar genetic susceptibility and predisposing markers in dog and human (MHC). Again, the similarity in MHC genes further suggests that results of investigating other genetic markers in dogs can have implications for humans as well [113].

With the complex MHC structure and multifactorial etiology of the traits, the causal variant(s) responsible for the effect are challenging to resolve and underlying pathogenic mechanisms difficult to understand. Furthermore, up to 90% of AS patients are positive for *HLA-B*27* [143], but only 5% of the individuals with the HLA risk allele will actually have the disease. Thus far at least twelve loci [i.e., *endoplasmic reticulum aminopeptidase 1* (*ERAP1*)] and chromosomes 2p15 and 21q22 have shown evidence of association with AS in Europeans [143,152,153,170].

2.3.8 THE EXAMPLE OF CELIAC DISEASE

One of the few well understood pathogenesis of HLA related diseases, is celiac disease, a chronic inflammatory disease, with an identified genetic component (DQ2 in cis or trans [(*DQA1*05-DQB1*02* with *HLA-DRB1*03:01*) or DQ8 (*HLA- DQA1*03:01-DQB1*03:02* with *HLA-DRB1*04*)] and a known environmental trigger (here, glutein) (Fig. 11)[124-126]. Celiac disease is an autoimmune-mediated systemic disorder characterised by a permanent intolerance to wheat gluten, primarily affecting the gastrointestinal tract [171]. In celiac disease, the MHC class II molecules present deamidated gluten particles to T cells in the small intestine, resulting in an abnormal CD4+ T-cell-initiated immune response to gluten. In addition, evolutionary studies have indicated that an interaction between at least two genes of the *HLA-DRB1*03:01–DQ2* haplotype is needed for disease predisposition [124]. Indeed, these observations of celiac disease may assist to identify the pathogenic mechanisms in other autoimmune and inflammatory disease.

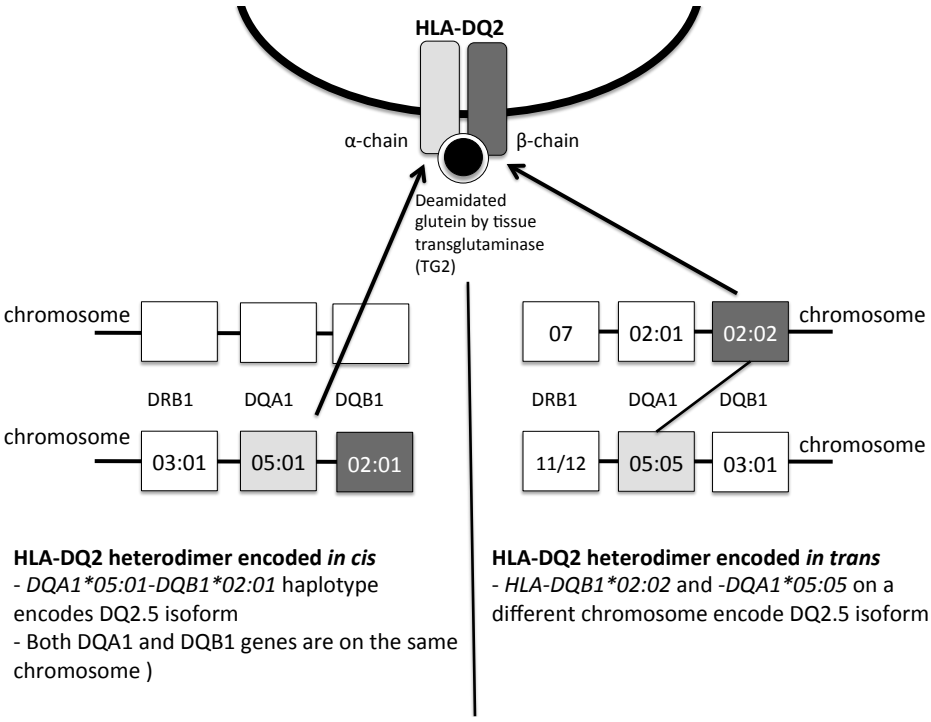


Figure 11 Most patients with celiac disease express a particular HLA-DQB1 heterodimer with *HLA-DQA1*05* and *HLA-DQB1*02* on the cell surface of APC, either encoded by genes that are carried in same chromosome (cis) or in different haplotype on different chromosome (trans) [124]. The MHC class II molecule presents gluten peptides that have been deamidated by tissue transglutaminase (TG2) to $CD4^+$ T cells[124].

2.3.9 MHC (NON-HLA) DISEASE ASSOCIATION WITH DISEASE

Several reports have associated various non-HLA gene polymorphisms, including *TNF,* lymphotoxin alfa *(LTA),* the human MHC class I chain-related gene A/B (*MICA/MICB*), *BTNL2, C4* and the receptor for advanced glycation endproducts (*AGER* or *RAGE*) with autoimmune and inflammatory traits [134,142].

Cytokine-coding genes, such as the tumor necrosis factor (*TNF*) produced by macrophages, and lymphotoxin alfa and beta (LTA and LTB) produced by lymphocytes, have key roles in the regulation of the inflammatory response and granulomatous inflammation [129,140,172]. The MHC region codes for three complement coding proteins; complement C2, C4 and complement factor B. The major function of the complement proteins is the activation of the complement pathway, which results in the formation of the membrane attack complex. Interestingly, the variation of the C4 gene numbers and the deficiency of C4 proteins (less than two copies of the gene) have both been associated with several autoimmune diseases [142]. The receptor for advanced glycation end products (*RAGE* or *AGER*) is a member of the immunoglobulin gene superfamily and *BTNL2* is a member of the butyrophillin family, both encoded by the MHC class III region. *RAGE* regulates autophagy and apoptosis and has been shown to associate with RA and sarcoidosis [173,174].

Furthermore, in sarcoidosis, RAGE is expressed in granulomas and subjects carrying the −374 T allele (rs1800624) have an increased RAGE expression [174]. BTNL2 acts as a co-stimulatory molecule in T-cell activation and evidence of disease association has been reported, e.g. with Grave's disease and sarcoidosis [173,175].

2.4 MHC CONTRIBUTES TO SARCOIDOSIS PREDISPOSITION

2.4.1 CLINICAL CHARACTERISTICS OF SARCOIDOSIS

Genetic factors play an essential role in immunodeficiency and the susceptibility to infectious and inflammatory diseases, including sarcoidosis (MIM: 181000). Sarcoidosis is a multi-organ immune-mediated disorder characterised by the presence of non-caseating epithelioid granulomas in affected organs [176]. Sarcoidosis, first recognized over 120 years ago, is a systemic disease of unknown aetiology and with a varying clinical course [4]. It is a complex disease, with interactive effects from the environment and multiple genes [177], and has heritability estimates (see chapter 2.1.1) as high as 66% [178] (i.e., significantly higher than in many autoimmune and inflammatory disorders). Previous studies suggest that sarcoidosis is recessively inherited disease with incomplete penetrance [179].

Sarcoidosis occurs throughout the world affecting all races but with differences in prevalence and incidence with respect to geographical locations [180]. Similar rates are observed in Northern Europe and the US. In Finland, on the basis of national study in 1984, the crude prevalence and annual incidence of sarcoidosis in 20 to 65-year-olds were estimated as 28.2 per 100000 and 11.4 per 100000, respectively [5,6]. Among African Americans, sarcoidosis is three times more common than in European Americans. The lowest rates are probably observed in Japan, with an incidence of 1-2/10^5 [5,6,181-183]. Sarcoidosis is uncommon among children and adolescents, and many studies have reported an increase of incidence during mid-adulthood [176,181]. The average mortality of sarcoidosis ranges from 1–6% [180,184,185].

The diagnosis of sarcoidosis is established when clinical and radiological observations (Fig. 12) are supported by histological evidence of noncaseating epithelioid cell granulomas (in a biopsy sample) and other diseases [e.g., tuberculosis, MS, Chronic beryllium disease (CBD] have been excluded [180]. In principle, all organs of the human body may be affected by sarcoidosis. Most often lesions are found in the lungs (90% of all patients) and bronchial walls, in lymph nodes, eyes, liver/spleen, skin, CNS, heart and the upper airways. Clinical manifestations depend on the location of the inflammation. Up to 50% of patients have extrapulmonary disease localisations (Fig. 13) [5,186-188]. A biopsy can be obtained from the lung parenchyma (a

transbronchial biopsy), a skin nodule or enlarged peripheral lymph node. The biopsy is difficult to obtain from patients with suspected neurosarcoidosis and/or cardiac disease, and these manifestations are typically under-diagnosed [189]. Common clinical manifestations for sarcoidosis include chest pain, uveitis, nodules in the skin, erythema nodosum, arthralgia, dyspnea and cough. [4]. Sarcoidosis can also stimulate unusual overproduction of vitamin D3 (calcitriol) by activated macrophages and granulomatous tissue causing a complication referred as hypercalcemia. In sarcoidosis hypercalcemia often be accompanied by renal insufficiency [190].

Five radiographic stages of pulmonary sarcoidosis

Stage 0	Normal chest X-ray
Stage I	Bilateral hilar lymphadenopathy (BHL)
Stage II	Parenchymal shadows and bilateral hilar lymphadenopathy in chest
Stage III	Parenchymal shadows alone
Stage IV	Fibrotic changes

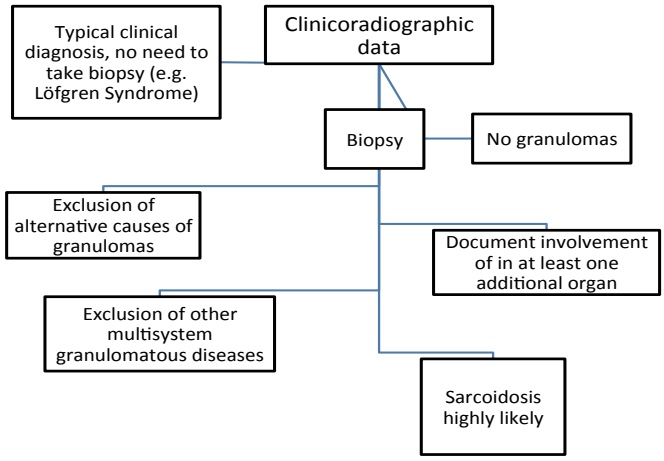


Figure 12 Radiological staging of pulmonary sarcoidosis and overview of diagnostic procedures. In addition to a typical clinical picture consistent with sarcoidosis, a confirming tissue biopsy is also required with the presence of noncaseating granulomas (Modified from [4,187])

Previous studies have indicated that sarcoidosis is not a single disease entity but a syndrome of heterogeneous diseases. The clinical picture and prognosis of sarcoidosis vary starting from spontaneous remission within 1-2 years to chronic disease with functional insufficiency of many organs. Indeed, the majority of the sarcoidosis patients with pulmonary involvement alone have a favourable prognosis but approximately 20% develop a chronic, disabling disease. The reason, why in some patients sarcoidosis resolves and in some becomes chronic, is poorly understood [191]. The resolving pulmonary sarcoidosis is typically associated with bilateral hilar lymphadenopathy (BHL) and non-fibrotic pulmonary infiltrates. Löfgren syndrome, the acute form of sarcoidosis accompanied with erythema

nodosum (EN) and acute arthritis and/or uveitis, is a genetically distinct subgroup of sarcoidosis with high frequency of HLA-DRB1*03 and has typically a favourable prognosis. Löfgren syndrome is common in Scandinavian patients, but uncommon in e.g., African-American and [192].

In general, an asymptomatic patient with only lymphadenopathy and/or with only slightly abnormal lung function can be followed up without the need for therapy. In progressive disease, when damage of organ function is suspected or already present, corticosteroids are the first line of therapy. In addition, TNF-inhibitors have been shown to improve lung function and reduce extra-pulmonary manifestations [193,194]. New therapies for sarcoidosis warrant a better understanding of the pathogenesis of sarcoidosis and the immunological pathways linked to the disease. In the end-stage pulmonary and cardiac sarcoidosis, organ transplantation can be considered. However, granulomas can re-occur in transplanted organs [195].

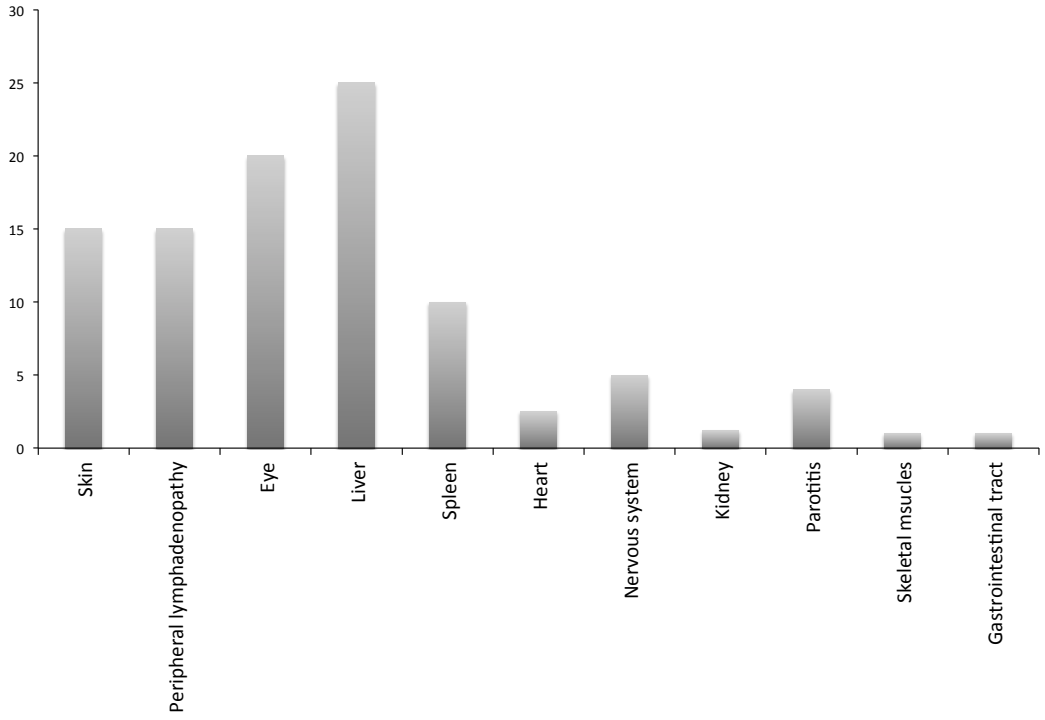


Figure 13 The average estimations of the extra-pulmonary manifestations of sarcoidosis (reviewed in [187]). In Finland (n=571, sarcoidosis patients seen at Mjölbolsta Hospital, Finland during years 1955-1987)[5], the prevalence of extra-pulmonary locations (>1%) sarcoidosis were the following: skin (5%), peripheral lymphadenopathy (25 %), eye (7%), spleen (4%), EN (erythema nodosum) (13%).

2.4.2 PATHOPHYSIOLOGY OF SARCOIDOSIS

The most probable pathophysiology of sarcoidosis is the dysregulation of the innate immune response to unidentified inhaled or infectious antigens [196,197]. To date, no single causative environmental trigger has been identified. However, increasing evidence suggests that particular disease

phenotypes have different environmental triggers including mold, mycobacterial or propionibacterial antigen (*Mycobacterium tuberculosis*, (*Propionibacterium acnes, P. granulosum,* respectively), inorganic particles, and insecticides [177,197-201].

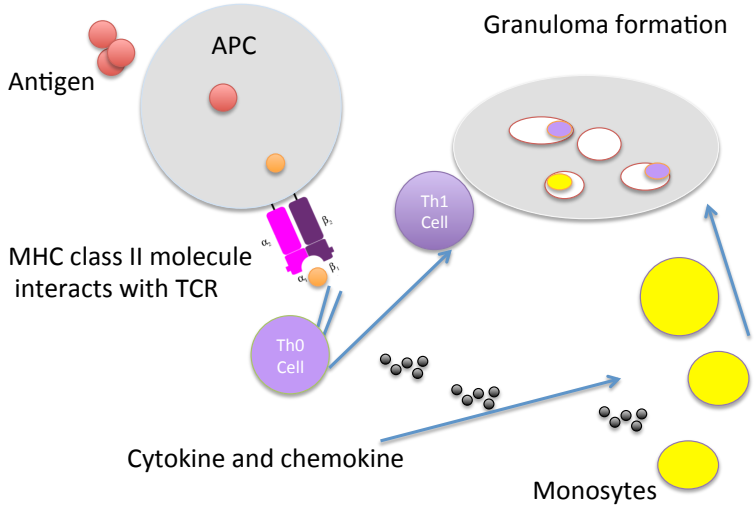


Figure 14 The development of nonceseating granulomas in sarcoidosis. Foreign antigen(s) are presented by antigen presenting cell (APC) to promote the release of cytokines (TNFα, IFNγ, IL-2) from monocytes and type 1 T helper cells (Th0). This causesthe T-effector cells (Th1) and monocytes to gather at the disease site. TCR=T cell receptor [202]

The immune-pathogenic mechanism of sarcoidosis is complex (Fig. 14) [162]. Sarcoidosis is characterized by granulomas, comprising of epithelioid cells, fibroblasts that are encircled by mononuclear cells and lymphocytes (CD4+ Th cells, CD8+ T cells and B cells). Granulomatous disorders include a wide spectrum of common and rare disorders, including chronic granulomatous disease (CGD), tuberculosis and Crohn's disease [202]. The formation of a granuloma is dependent on the response to antigenic stimulation and presence of previously mentioned cells [203]. The aggregation of cells in granulomas is probably promoted by the immune response to the pathogen [158]. Based on immunohistochemical staining of granulomas, most of the affected lymphocytes are CD4+ T cells, but CD8+ T cells are also observed in the periphery of the granulomas [204,205]. In addition, tissues affected by sarcoidosis have unregulated expression of interferon (IFN)-γ, TNF, and interleukins (IL-2, IL-12, IL-18 and IL-27), which regulate the CD4+ T lymphocytes to differentiate and activate the macrophage response [191,206]. Patients with sarcoidosis also have an increased proportion of T cells in the bronchoalveolar lavage fluid [207]. Evidence of association between sarcoidosis and the decrease of T regulatory cells, the negative immunological feedback signals, and the natural killer cells highlight the importance of the immune system in the disease predisposition [191,208,209].

2.4.3 GENETIC FACTORS OF SARCOIDOSIS

Epidemiological findings in family, population-specific and multi-ethnic studies suggest that sarcoidosis is partially a genetic disorder, and that genetic factors may determine the clinical course [178,182,210]. Both linkage and association studies, including the recent GWASs, have shown a strong role for MHC in susceptibility to sarcoidosis [7]. Associations with the particular MHC genes (e.g., *HLA-DRB1*) have repeatedly been encountered in various ethnic groups and disease severities. As a complex disease, effects from multiple genes and environmental factors are needed for disease predisposition. Many of thee effects are probably relatively small, except in the case of the HLA genes (with OR for predisposing genes ranging between 2 and 12 [211]), which have been characterized as the main contributor in disease susceptibility to sarcoidosis across different ethnic populations [212].

2.4.3.1 MHC and sarcoidosis

The HLA genes are the most extensively studied susceptibility genes in sarcoidosis (Table 4) [164]. Both predisposing and protective HLA class I and class II associations have been found in different ethnic groups: some HLA risk loci are shared, whereas others are ethnicity-specific [211]. The current understanding of the pathophysiology of sarcoidosis suggests that HLA class II genes, mostly the *HLA-DRB1* alleles, are likely to be involved in sarcoidosis susceptibility. However, some studies suggest that HLA class I and class II associations have independent effects from both loci [213]. Typically, alleles *HLA-DRB1*03, *11, *12, *14* and **15* represent the risk factors and *HLA-DRB1*01* and **04* are protective ones. In addition to allelic association, the peptide-binding sites of the HLA genes have been indicated to be critical in disease pathogenesis [214,215]. The published HLA associations with sarcoidosis are summarized in Table 3.

Despite several clinical and genetic studies, there are no biomarkers that can estimate the prognosis or evaluate the treatment effects in sarcoidosis. The only exception is that *HLA-DRB1*03* positive patients with Löfgren syndrome (LS) have a favourable prognosis compared with *HLA-DRB1*03* negative LS patients (95% vs. 49%, respectively) [192]. In a Swedish study, sarcoidosis patients with peptide-specific T cells and *HLA-DRB1*03:01* had a favourable prognosis and the antigen-driven stimulation of the T cells was supported [216]. Before using HLA types as genetic markers for an association with a particular disease phenotype, a replication of the variant is needed in independent populations. However, it should be noted that chronic beryllium disease (CBD), a sarcoidosis-like granulomatous disease, has strong HLA associations showing an increased prevalence of HLA-DPB1 alleles containing a glutamic acid at amino acid position 69 (Glu69) with ORs over 10 [217,218]. Because of similar clinical pictures in sarcoidosis and CBD, it has been hypothesized that chronic sarcoidosis could have the same risk factor as Glu69 (typically HLA-DPB1*02:01) [219]. However, previous studies have shown conflicting results regarding sarcoidosis and HLA-DPB1 associations

[220]. Future studies are needed to address the HLA-DPB1 interaction with other MHC genes in sarcoidosis, especially in the development of chronic sarcoidosis.

Table 3. *HLA associations with sarcoidosis and its various disease phenotypes*

HLA association	Susceptibility/OR	Prognosis	Subphenotype	Reference
HLA-A*03, -B*07	Risk factor	Persistent		[213,221]
HLA-A*01, -B*08	Risk factor	Good prognosis	Arthritis	[162,164,213,222]
HLA-DRB1*01	0.1-0.5			[211,222-224]
HLA-DRB1*03 (HLA-DQB1*02:01)	1.9 (all sarcoidosis) / 6.7-12.5 (LS)	Good prognosis	Acute onset, LS	[162,211,224]
HLA-DRB1*04	0.2-0.7 / 7.5		Ocular sarcoidosis, Heerfordt's syndrome	[211,225]
HLA-DRB1*08:03	2.0		Ocular sarcoidosis, Japanese patients	[226]
HLA-DRB*11:01	1.5-2.0		Stage II/III chest X-ray	[214,224]
HLA-DRB1*12:01	2.1-14.7			[211,214]
HLA-DRB1*14	1.8	Persistent	Lung-predominant sarcoidosis	[163,227,228]
HLA-DRB1*15:01 (HLA-DQBQ*06:02)	1.3-1.8	Persistent		[214,224,229]
pocket 4 (DRAla71 in *HLA-DRB1*), pocket 9 (DQPhe9 in *HLA-DQB1*)	2.2 and 2.4			[215]

LS, Löfgren syndrome

2.4.3.2 Non-HLA and sarcoidosis

From the MHC class III region, most importantly, the variants in TNF and *BTNL2* genes have shown evidence of association with particular disease phenotypes of sarcoidosis [175,230]. TNF is a key factor in granuloma formation and it is highly expressed in sarcoidosis patients. TNF has been shown to associate especially with progressive disease [231]. Indeed, the most promising non-HLA association with sarcoidosis is for a variant in the *BTNL2* gene (missense SNP rs2076530)[175]. It has been indicated that the non-functional BTNL2 protein, which is structurally similar with CD80 and CD86 molecules, acts as a negative costimulatory molecule [232]. In theory, BTNL2 could increase T lymphocyte activation, which is compatible with the pathophysiology of sarcoidosis [175]. The rs2076530 association with sarcoidosis has been replicated in many populations (e,g., [230]), but contradictory findings exist as well, especially regarding the strong LD between BTNL2 and HLA class II alleles [233,234]. BTNL2 is also associated with numerous autoimmune and inflammatory diseases, including type 1

diabetes (T1D) and Crohn's disease [235,236]. In addition, also other SNPs in BTNL2 has been shown to associate with sarcoidosis e.g., rs9268480 [237].

However, not many of the found non-HLA associations with sarcoidosis have been replicated in an independent sample: either the markers show contradictory results (e.g., [238,239]) or await replication (e.g., Rab23, CCR5 and IL23R [240]). Again, the strong LD existing in the MHC makes it extremely complex to resolve which gene represents the primary association. The functionality of a particular variant is often not defined, as the strong LD within the MHC genes complicates finding the causal variant [212,241]. More extensive studies across MHC (including the non-HLA genes) are needed in order to describe the extended MHC haplotypes accurately. Using larger samples, genotyping multiple genes and using proper statistical approaches may assist us to identify the causal variants.

2.4.3.3 GWAS and sarcoidosis

A total of six GWASs have been carried out to investigate the genetic susceptibility to sarcoidosis. With traditional candidate gene case-control approaches, the studies have pointed out novel non-HLA candidate regions of interest (Tables 3 and 4). One of the most promising non-HLA markers is *annexin A11* (*ANXA11* on 10q22.3) [242] that encodes a molecule involved in various functions in calcium signalling, apoptosis and cellular proliferation. ANXA11 among HLA class II molecule, are detectable in human B-cell exosomes [243], which are involved in regulation the immune response [242,244]. The association of sarcoidosis with the non-synonymous SNP (rs1049550; $P = 1.0 \times 10^{-12}$, OR=0.62) is significant across different populations. The association has been independently replicated in two European samples [237,245,246] and in a meta-analysis of both European Americans and African Americans [244]. Furthermore, the most recent GWAS [244] identified a novel SNP-SNP interaction between *ANXA11* (rs1049550) and the HLA (rs9268839), which sheds light on the importance of the exosomes in sarcoidosis predisposition. Indeed, the suggested interaction between *ANXA11* and HLA warrants replication studies, and it is tempting to speculate that in case of sarcoidosis, the dysregulation of the immune response (via HLA) may result in the dysfunction of *ANXA11,* that could in turn hamper apoptosis of activated inflammatory cells [244].

In addition, previous GWASs in sarcoidosis have indicated other interesting non-HLA susceptibility loci at 6p21.32 (rs715299 in *NOTCH4*) [237], 10p12.2 (rs1398024 in *C10orf67*) [247], 11q13.1 (rs479777 in *CCDC88B*) [251], 12q13.3–q14.1 (rs1050045 in the 3'-UTR of *OS9*) [250]. However, in many GWASs the potential functional relevance of the detected associations remains unknown.

Table 4. *Non-HLA gene association with sarcoidosis.*

Variant	Position	Yes	P value/ OR	Reference
AGER (rs1800624)	6p21		0.004/ 2.8	[174]
ANXA11 (rs1049550)	10q22.3	Yes	1 x 10^{-12} / 0.62	[242,244]
BTNL2 (rs2076530)	6p21		1.1 x10^{-8}/1.6-2.75	[175]
C10ORF67 (rs1398024)	10p12.2	Yes	4.24 x 10^{-6} / 0.81	[247]
NOTCH4 (rs715299)	6p21.3	Yes	6.51 x 10^{-10} (meta-analysis)	[237,248]
TNF (e.g. rs1800629)	6p21.3		>0.001/ 0.43 *	[172,216]
BAG2, C6orf65, KIAA1586, ZNF451 and RAB23 (rs10484410)	6p12	Yes	2.64 x 10^{-4}/ 1.72	[249]
OS9(rs1050045)	12q13.3–q14.1	Yes	9.22 x 10^{-8}/ 1.24	[250]
CCDC88B (rs479777)	11q13.1	Yes	2.68 x 10^{-18}/ 1.18	[240]

ACE, Angiotensin-converting enzyme; AGER, advanced glycation end product receptor; ANXA11, Annexin A11; BAG2, BCL2-associated athanogene 2; BTNL2, Butyrophilin-like 2 (MHC class II associated); CCDC88B, coiled-coil domain containing 88B; C6orf65, chromosome 6 open reading frame 65; C10orf67, chromosome 10 open reading frame 67; OS9, osteosarcoma amplified 9; NOTCH4, Notch homolog protein 4 ; RAB23, RAB23, member RAS oncogene family; TNF, tumor necrosis factor; ZNF451,zinc finger protein 451;

*rs2800629 associated with the good prognosis.

2.4.3.4 The challenges in sarcoidosis genetic studies and future directions

The challenging problem with complex genetic studies is to find the genes responsible for the trait or disease. As there is no well-established animal model for sarcoidosis, collaborative studies are needed to obtain sufficiently large samples sizes. Screening of the possible candidate region is typically done via genome-wide studies (e.g. GWAS), using samples from families, sib-pairs, or case-control collections (unrelated affected individuals and unrelated unaffected individuals). As sarcoidosis is a rare disease, the collection of large families for linkage studies or affected subjects for GWAS is not cost-effective. Thus, to date, many of the published genetic sarcoidosis studies are traditional candidate gene approaches that do not detect novel susceptibility loci. In addition, it has been shown that especially the GWASs are strongly influenced by the selected study population, and the population stratification remains a challenge in genetic studies of sarcoidosis [142,233,237]. The impact of new research advances (e.g., NGS) with sarcoidosis remains rather modest so far. To overcome the strong LD in MHC and to detect independent susceptibility variants, more improved analytical methods (e.g., using extended haplotyping or RNA-based studies) and larger samples sizes with well-established diseases phenotypes are needed.

3 AIMS OF THE STUDY

The overall objectives of this study were to investigate the polymorphic profile of the MHC region in Finnish subjects and to address practical issues that should be taken into account in studying MHC disease association. In addition, the study aimed to investigate the MHC as susceptibility loci for sarcoidosis, a complex immune-mediated disease. The purpose of these activities was to find gene variants that predispose patients to sarcoidosis and predict the prognosis of sarcoidosis. If a significant correlation between the course of sarcoidosis and a genetic marker is found (i.e., disease phenotype associated marker), this may have clinical benefits in the treatment and the follow-up of the sarcoidosis patients.

Our specific aims were:

- To characterize the MHC profile among Finnish study subjects using extended MHC haplotypes covering the MHC class I, II and III region
    - To estimate the possible implications of the extended MHC haplotype to the study of the MHC related diseases (I)
    - To present practical guidelines and highlight challenges in MHC association studies (I)
- To evaluate several different levels of the MHC association with sarcoidosis and its prognosis
    - To estimate whether the extended MHC haplotypes are advantageous in disease association studies in order to overcome the strong LD in the MHC (II, III, IV)
    - To estimate whether the previously identified MHC markers are associated with the Finnish sarcoidosis patients (II, III)
    - To detect novel MHC susceptibility variants for sarcoidosis phenotypes that are shared among Dutch, Czech, Swedish and Finnish sarcoidosis subjects (IV)
    - To pinpoint MHC disease variants associated with the disease course of sarcoidosis in order to distinguish the disease phenotype with genetic marker(s) (II, III, IV)

4 MATERIALS AND METHODS

4.1 STUDY COHORTS

4.1.1 FINNISH POPULATION SAMPLE (VITA) (I, II, III, IV)

The "VITA" population sample consisted of 150 consecutive voluntary subjects coming to Vita Laboratory Ltd, Helsinki, Finland for a health survey before accepting a new occupational post [252]. Because of the recruitment setting, subjects were unlikely to be related. Briefly, the mean age of the subjects was 33.7 years (range 18-60), and 67% were females. A more detailed description of the sample is presented in Seppänen et al., [252]. The sample was used in our sarcoidosis study as a control group representing a healthy Finnish sample. However, we cannot reject the possibility that subjects who have later developed sarcoidosis are included in this sample. This study was approved by the Ethics Committee of Department of Medicine, Hospital District of Helsinki and Uusimaa, Helsinki, Finland (approval Dnro 6/E5/2001, 25.1.2001). All study subjects provided a written informed consent.

4.1.2 FINNISH SARCOIDOSIS SAMPLES (II, III)

A total of 188 Finnish sarcoidosis patients with pulmonary sarcoidosis and 150 control subjects were included in studies II and III (Table 6). The sarcoidosis patients were recruited from 17 pulmonary units in Finland. In all cases the diagnosis was based on the clinical presentation and was supported by a positive tissue biopsy and/or laboratory tests. Fifty percent of the patients had previously participated in a controlled clinical trial with a 5-year follow-up [253]. In this study, patients with a clinical presentation likely to have a favourable clinical course, i.e. patients with LS, and patients requiring immediate treatment, were purposely excluded. As these patients would have been underrepresented in the current study, the sample set was expanded with additional patients having an acute onset sarcoidosis and patients with chronic sarcoidosis. The sarcoidosis patients underwent the following tests: chest radiograph, lung function tests (spirometry, diffusion capacity), electrocardiography (ECG) and blood samples [blood picture, liver enzymes tests, serum calcium and creatinine, serum lysozyme, serum angiotensin-converting enzyme (S-ACE)]. All patients had pulmonary sarcoidosis and 59 had one or more extra-pulmonary lesions. All the patients had a robust diagnosis of sarcoidosis and had a follow-up of at least 5 years. The Finnish Ministry of Social Affairs and Health approved the request to study hospital records of the patients (approval Dnro 362/E5/05). All study subjects provided a written informed consent.

4.1.3 THE SARCOIDOSIS COLLABORATION STUDY -SAMPLES (IV)

Collaboration with three European Sarcoidosis Research groups increased the sample size in Study IV (Table 5) [224,254-256]. All the patients had pulmonary sarcoidosis. The Swedish sample consisted of 219 sarcoidosis patients and 360 healthy controls. The Swedish study was approved by the Ethics Committee of Karolinska Institute forskningsetikkommitté Nord (Dnr 02-438: Dnr 2007/340-32), Stockholm, Sweden. The Czech sample consisted of 218 sarcoidosis patients and 180 healthy controls, and this study was approved by the Ethics Committee of Faculty Medicine of Palacky University and Faculty Hospital Olomouc (IGAMZCR NT/11117), Olomouc, Czech Republic. The Dutch sample consisted of 180 sarcoidosis patients and 180 healthy controls. The study was approved by the Ethics Committee of the St. Antonius Hospital, Nieuwegein, The Netherlands. All participants gave a written informed consent.

Table 5. *Patients and controls included in the study*

	Sarcoidosis		Controls	
	Before QC	After QC	Before QC	After QC
Discovery Sample				
Finland	188	187	150	150#
Replication Sample				
Sweden	219	190	360	358
Dutch	180	180	180	173
Czech	218	208	180	178##
Total	805	765	870	859

QC=quality control, detailed in methods chapter 4.3; Plink software [257]
#149 controls used in the C4 studies
179 controls used in the haplotype analysis

4.1.4 SARCOIDOSIS PHENOTYPES

Patients were further divided into those with a disease resolved within 2-4 years and those with persisting activity at that time point [163]. The activity of the disease was based on chest radiographic findings, lung function tests and laboratory values (S-ACE activity and lysozyme concentrations). The characteristics of the Finnish, Swedish, Dutch and Czech sarcoidosis patients are described in detail in Table 6.

Table 6. *Patients were subgrouped according to LS and the disease activity.*

	NL	LS	NLR	NLP	NL, no subgroup info available
Discovery Sample					
Finland	168	19	79	89	-
Replication Sample					
Sweden	112	78	33	75	4
Dutch	180	0	90	90	-
Czech	169	39	47	83	39
Total	629	136	249	337	43

NL, Non-Löfgren; LS, Löfgren; NLR, non-Löfgren syndrome and resolved disease; NLP, non Löfgren and persistent disease

4.2 GENOTYPING METHODS

4.2.1 DNA EXTRACTION

The DNA of the Finnish samples was extracted from the buffy coat fraction of whole blood by using NucleoSpin ® QuickPure columns (Macherey-Nagel, GmbH & Co. KG, Düren, Germany) or Gentra Puregene Kit (Qiagen, Vienna Austria; according to the Manufacturer's protocol.

4.2.2 HLA TYPING (I, II, III, IV)

HLA genotyping was performed and/or analysed in an EFI (European Federation for Immunogenetics) accredited HLA Laboratory at the Haartman Institute, part of the University of Helsinki, Finland. Commercial HLA kits were used primarily (except *HLA-DQB1* typings), and the reactions were performed according to the Manufacturers' instructions and using the HLA nomenclature, release 3.5.0 (IMGT/HLA database). In principal, alleles that were identical in exons 2 and 3 (MHC class I) or exon 2 (MHC class II) were not resolved; expressed alleles in this category share the amino acid sequence of their antigen binding grooves. Other ambiguities were resolved with high resolution SSPs or sequencings HARPs. Two different people carefully interpreted all the HLA genotypes. The pipeline used for the HLA typing is presented in Fig. 15.

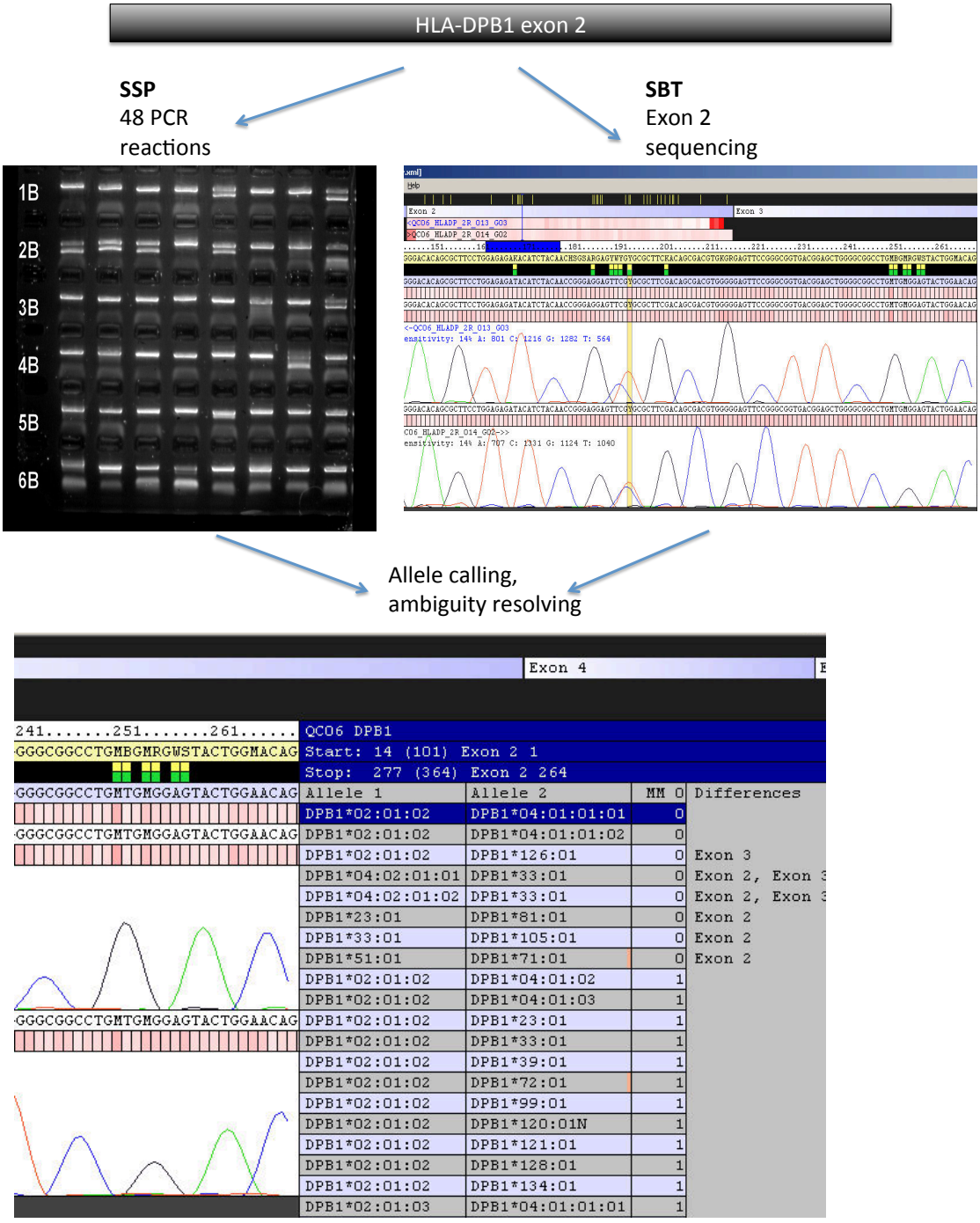


Figure 15 Pipeline for HLA genotyping with SSP or SBT using MHC class II gene, HLA-DPB1 as an example

SSPs (Olerup SSP AB, Stockholm, Sweden) were used for genotyping of the *HLA-A* and *-B* genes (Study I). PCR reactions from agarose-gel were evaluated manually and the alleles were called with SCORE software. SBT (Invitrogen™, Life Technologies, Carlsbad, CA, USA) was used for detecting *HLA-DRB1* alleles (Study I, II, III and IV). The sequencing reactions products were

identified with ABI Prism 3100 capillary electrophoresis (PE Applied Biosystems, Foster City, CA) and sequence interpretation was performed using the Assign-SBT v3.5+ software (Conexio Genomics, Applecross, Australia) and evaluated manually. The genotyping of *HLA-DQB1* (Study I) was performed with a panel of lanthanide-labelled oligonucleotide probes, as described previously. For genotyping the *HLA-DPB1* gene (Study I and III), both SSP and SBT were utilized, and the reactions performed and alleles called as previously presented (Fig. 14). As a quality control procedure, the *HLA-DRB1* and *HLA-DQB1* alleles were matched with the known LD pattern. For Study IV, we received *HLA-DRB1* low-resolution genotypes for Swedish sample (NL+L n=184, controls n=187) and Czech Sample (NL+L n=90, controls n=179).

4.2.3 C4 COPY NUMBER VARIATION AND C4 ALLOTYPING (I, II)

The complement C4A and C4B gene copy numbers and a 2-bp silencing insertion at exon 29 (C4A CTins) in the C4A gene were analysed by quantitative genomic realtime-PCR (qPCR or RT-PCR) Rotor-Gene 3000 or 6000 (Corbett Research, Sydney, Australia) using SYBR® Green labeling (ABsolute™ qPCR SYBR® Green Mix, AB-1159, Bgene, Epsom, UK) and Brilliant SYBR® Green QPRC Master Mix (Staratagene, AHDiagnostics, Skärholmen, Sweden). Data was analysed with the Rotor Gene software v 6.0 (Qiagen) according to the Manufacturer's instructions and using the amplification of a housekeeping gene (beta-actin) to assess the sample's concentration. Detailed information on the qPCR protocols and primer sequences of *C4, C4B* and for *C4* CTins is given in Paakkanen et al [258]. The allotypes of C4 were resolved from serum samples by immunofixation [Carboxypeptidase B (Roche Diagnostics Gmbh, Mannheim, Germany) and neuraminidase (Type IV Sigma- Aldrich Chemie Gmbh, Steinheim, Germany), polyclonal anti-C4 antibody (DiaSorin Inc., Stillwater, MN)] and the samples were separated by electrophoresis as described previously. The *C4* CNV genotyping and allotyping were performed at the Transplantation Laboratory of the Haartman Institute.

4.2.4 SEQUENOM® GENOTYPING (I, III, IV)

Using a functional candidate gene approach, polymorphism of the selected genes from the MHC region, *AGER* (Study IV), *LTA* (Study I and IV), *TNF* (Study I and IV), *BTNL2* (Study I, III and IV) and *HLA-DRA* (Study I and IV), were genotyped with SNPs. All the selected genes were covered entirely including the 5' and 3'- flanking regions. The SNPs were selected from two publicly available databased, the HapMap database (http://hapmap.ncbi.nlm.nih.gov) and the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP). In the SNP selection, the validation status, tagging quality, minor allele frequency (>0.01) and gene structure were used to evaluate the SNPs.

SNP genotyping was performed based on the differences of the single base extension (SBE) products using the Sequenom MassArray iPLEX system with 9-10 ng of DNA as a template (Sequenom®, San Diego, CA, USA) with default settings. The multiplex PCR assays were designed with the AssayDesigner (Sequenom®, San Diego, CA, USA). The SNP alleles were called with MassARRAY® Analyzer and evaluated visually. We applied the following quality control filters: Assays (or samples) with a minimum total per-sample call rate of 90 %, SNP minor allele frequency (MAF) > 0.01, and Hardy-Weinberg equilibrium (HWE) > 0.001 or water contamination were excluded. The average success rate of the assays was 99 %. No discrepancies were observed. The SNP genotyping was performed at the Institute for Molecular Medicine Finland (FIMM) Technology Centre, Helsinki, Finland.

4.3 STATISTICAL METHODS

We followed the STREIS principles to immunogenomic data analysis [259]. In general, the statistical analyses were performed with PAWStatistics version 18 (PAWS, Inc., Chicago, IL, USA) if not otherwise stated. Allele, phenotype, haplotype, and amino acid frequencies were obtained by direct counting. The phenotype (positivity or a carrier) of a specific variant was considered when the subject had one or two copies of the variant. The phenotype describes the prevalence of the marker in a given sample/population (f =positivity of the variant/number of individuals in the sample). Each locus was tested for HWE using different software packages depending on the nature of the studied loci: ARLEQUIN 3.11[260], GENEPOP 4.1.4 [261] or Haploview [262] (biallelic SNPs only) was used to carry out Hardy-Weinberg testing. Marker distributions were statistically analysed with a Chi-square test ($\chi 2$) (or Fisher's exact test when appropriate), as well as with logistic regression (a stepwise forward logistic regression). In the stepwise logistic regression, variants were removed until only significant variants were left. If only SNPs were studied, the software package PLINK [257] was used for the $\chi 2$ analysis. A two-sided p-value less than 0.05 was considered statistically significant and ORs with 95% confidence intervals (CIs) were calculated as association measures. The results are shown as uncorrected p-values if not otherwise stated. When the values were corrected (Pc) for multiple comparisons, the Bonferroni method was used (Study II).

Multi-locus and multi-allele haplotype frequencies were estimated from allele data using the Bayesian method with PHASE v. 2.1.1 [263]. Missing values were excluded from the analyses. SNP haplotypes were constructed using Haploview [262]. Considering the small sample size, in order to exclude unreliable haplotypes, only haplotypes greater than 1% (observed more than 3 times) were used in the analyses. In Study III, the alleles with a too low observed frequency were lumped together [264,265]. To study the similarity of SNP haplotypes based on the genetic distance, the R-package 'ape' was used to create phylogenetic trees with Neighbour-joining algorithm according to

the method of Saitou and Nei [266]. LD between alleles of each pair of loci was tested. The strength of LD was quantified by LD measures (D' and r^2) and determined using either the Haploview software [262] (for biallelic markers) or the ARLEQUIN 3.11 software [260] (for multiallelic markers). Due to the strong LD within MHC, a logistic regression was performed to study the effect of each marker in relation to the other risk alleles, and to adjust the markers for particular covariates, e.g., sex. For *HLA-DRB1* and -*DPB1* genes, amino acid (aa) residues in polymorphic hypervariable regions (HVR) were derived from nucleotide sequence of exon 2 using [267] SKDM HLA tool (Study II, III, unpublished data). Software package PHASE [263] was used for estimating recombination rates between the selected MHC loci (Study IV). For illustrating the overall recombination hotspot in the MHC region the HapMap data was used.

In the Sarcoidosis collaboration study (Study IV), replication and meta-analysis of selected SNPs was performed from the Swedish, Dutch and Czech samples. Results from primary $\chi 2$ association analysis (using the discovery and replications sample sets) were combined for the meta-analysis using a random effects model (PLINK software) [257]. The I^2 measure was used to evaluate heterogeneity of the results between studies. The possible effect of population stratification on the results was studied more in detail using the Cochran-Mantel-Haenszel test [PLINK software] [257].

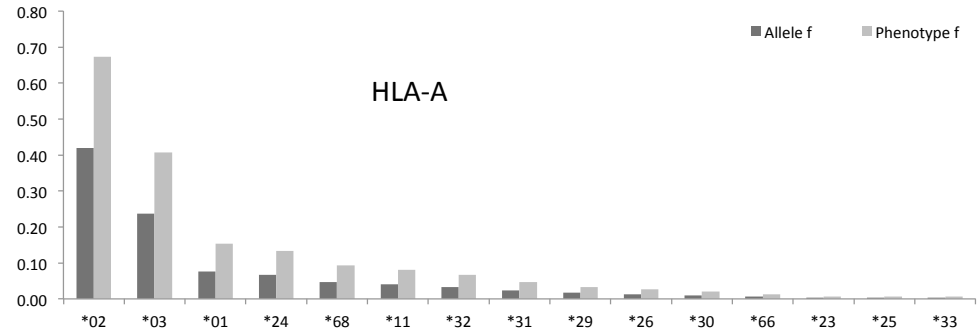
5 RESULTS

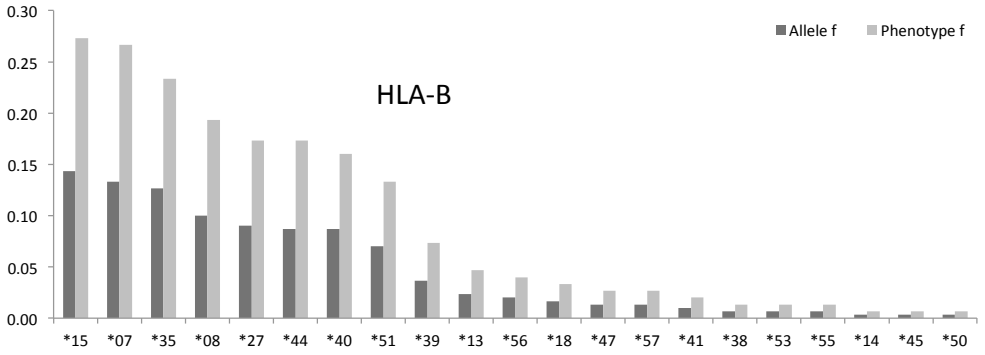
5.1 THE MHC PROFILE OF THE FINNISH SAMPLE (I)

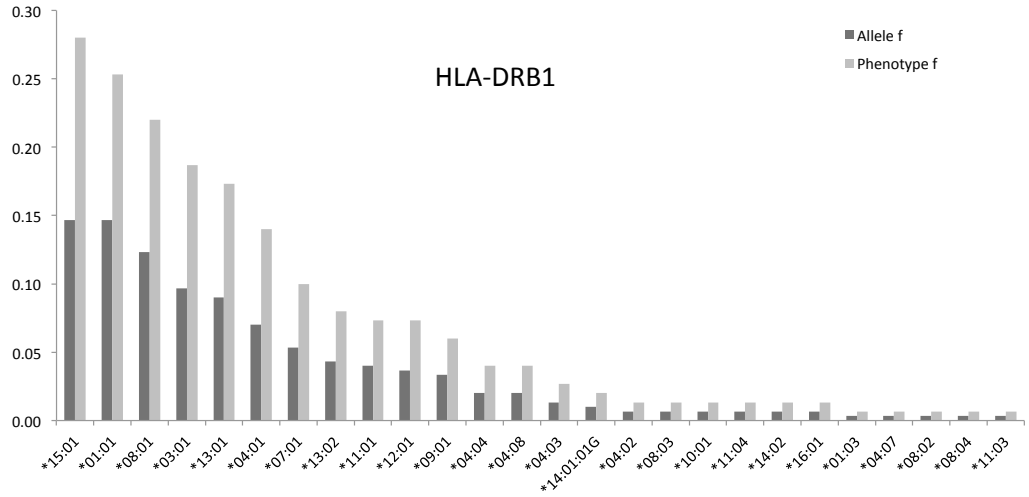
MHC alleles and haplotypes have differential distributions in different populations [2,108]. These distributions can be used to study genetic similarities between populations, and infer population history and evolutionary aspects. To explore the MHC profile of our population-based Finnish samples (n=150), we genotyped HLA class I (*HLA-A, -B*), class II (*HLA-DRA, -DRB1, -DQB1, -DPB1*), and a selected group of class III genes related to immunological pathways. We estimated the multi-locus extended haplotypes (see Fig. 8) and pairwise LD between loci to identify the most common MHC haplotypes and the LD patterns between tested variants.

5.1.1 HLA ALLELE DISTRIBUTIONS

The distribution of HLA alleles and phenotypes (i.e., carrier status for a particular allele) for the HLA-A, -B, -DRB1, -DQB1 and -DPB1 genes are shown in Fig. 16. No deviations from expected HWE proportions were observed. The allelic diversity was high at each HLA locus, with 15 HLA-A alleles, 21 HLA-B alleles, 26 HLA-DRB1 alleles, 11 HLA-DQB1 alleles and 19 HLA-DPB1 alleles measured. HLA-A, -B and –DQB1 alleles were detected at low-resolution level, explaining the lower amount of alleles compared with the high-resolution typed ones. Two HLA-A alleles (*02 and *03) and four HLA-B alleles (*15, *07, *35, *08) were observed at frequencies greater than 10%, and represented 66% and 50% of the allelic diversity observed at this locus, respectively. In MHC class II, four HLA-DRB1 (*15:01, *01:01, *08:01 and *03:01), five HLA-DQB1 (*06:02, *05:01, *02, *03:01, *04) and four HLA-DPB1 alleles (*04:01, *04:02, *02:01, *03:01) were observed with frequencies above 0.1, and represented 43%, 72% and 85% of the HLA class II diversity in Finnish population, respectively.


HLA-A
Allele f
Phenotype f
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
*02
*03
*01
*24
*68
*11
*32
*31
*29
*26
*30
*66
*23
*25
*33


HLA-B
Allele f
Phenotype f
0.30
0.25
0.20
0.15
0.10
0.05
0.00
*15
*07
*35
*08
*27
*44
*40
*51
*39
*13
*56
*18
*47
*57
*41
*38
*53
*55
*14
*45
*50


HLA-DRB1
Allele f
Phenotype f
0.30
0.25
0.20
0.15
0.10
0.05
0.00
*15:01
*01:01
*08:01
*03:01
*13:01
*04:01
*07:01
*13:02
*11:01
*12:01
*09:01
*04:04
*04:08
*04:03
*14:01:01G
*04:02
*08:03
*10:01
*11:04
*14:02
*16:01
*01:03
*04:07
*08:02
*08:04
*11:03

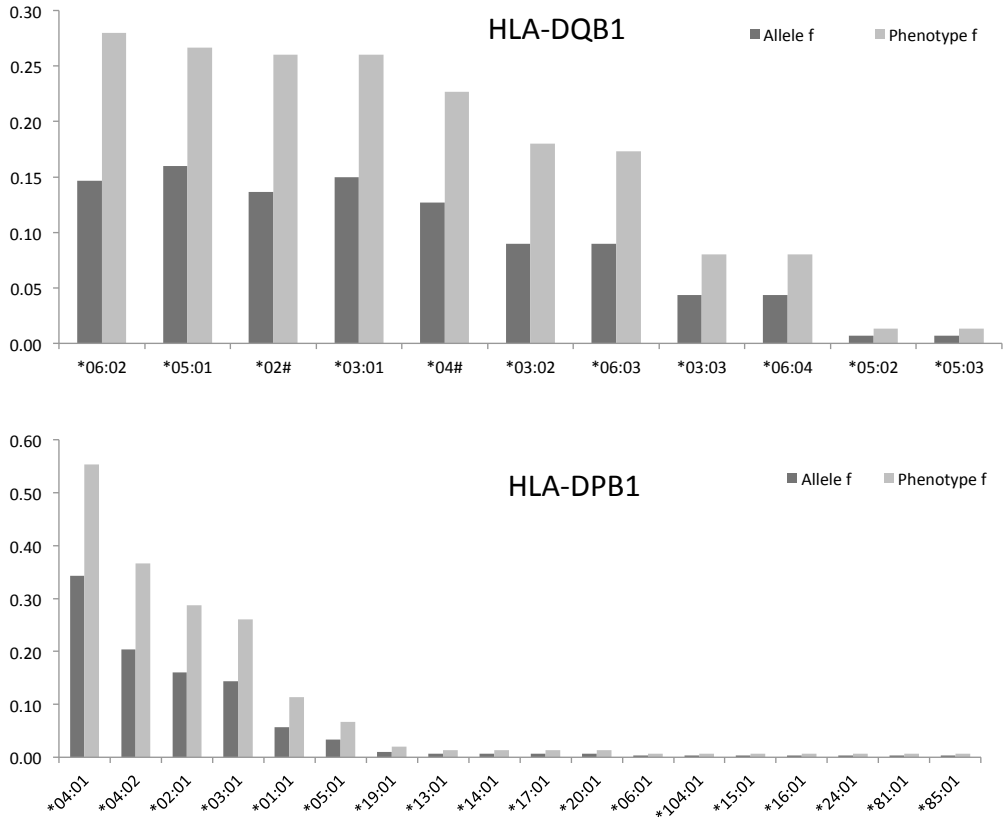


Figure 16 Traditional HLA genes and their HLA allele and phenotype (i.e., the carrier status for an allele) frequencies (%) in the Finnish sample (n=150) (I)

5.1.2 THE COMMON FINNISH HLA HAPLOTYPES

Several extended haplotypes with different gene combinations (two-locus, three-locus and five-locus) were analysed in the Finnish sample. The four most frequent Finnish *A~B~DR* –haplotypes were *A*03-B*35-DRB1*01:01* (7.1 %), *A*01-B*08-DRB1*03:01* (4.0%), *A*02-B*07, DRB1*15:01* (3.5 %) and *A*2-B*27-DRB1*08:01* (2.7 %). The most common Finnish *A~B~DR* –haplotype is uncommon elsewhere in Europe, except among Swedish Sami and in Russia (Fig. 17) [114,268-271].

The MHC class II haplotype analysis including *HLA-DRB1, -DQB1* and *–DPB1* alleles, indicated that while the *DR-DQ* block was constant (in strong LD), it showed polymorphism within the *HLA-DPB1* gene (Fig. 18). In addition, the four most common *HLA-DRB1* alleles, *HLA-DRB1*01:01, *15:01, *08:01* and **03:01* all existed with at least three different *HLA-DPB1* alleles (*HLA-DPB1*02:01, *04:01 and *04:02; HLA-DPB1*04:01, *04:02 and *05:01; HLA-DPB1*03:01, *04:01 and *04:02; HLA-DPB1*01:01, *04:01* and **04:01*, respectively).

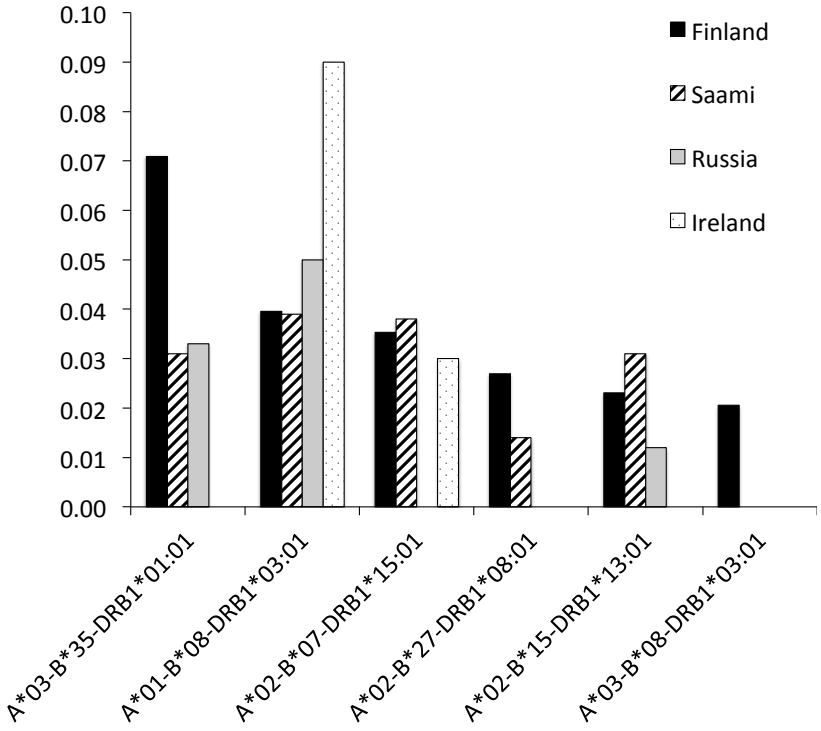


Figure 17 Haplotype *A~B~DR* frequencies in Finland, among Swedish Sami, in Russia and Ireland show distinct population differences on the haplotype frequency (I).

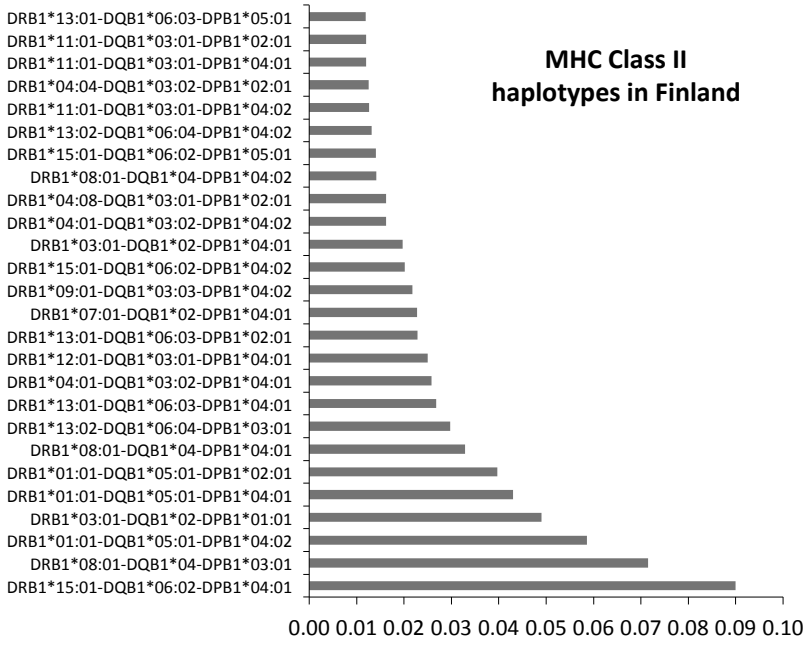


Figure 18 The most common (observed n > 3) MHC class II haplotypes in the Finnish sample

5.1.3 PAIRWISE LD BETWEEN HLA GENES

High LD between loci and population-specific recombination hotspots characterize the MHC region [272]. The association between the Finnish HLA loci was estimated with the global LD (unpublished results; Table 7) showing that the HLA alleles were in strong LD with one another; the only exceptions were *HLA-A* with *HLA-DRB1* and *-DPB1* and *HLA-DRB1* with *-DPB1*, neither of whichreached the significance level ($p<0.05$). Analysis of the HLA class II haplotypes and LD patterns show that, as expected, the linkage of *HLA-DRB1* to *HLA-DQB1* was stronger than to *HLA-DPB1*. This indicates that in relation to *HLA-DRB1*, several different *HLA-DPB1* alleles are found in the extended haplotypes, as shown in Fig. 18. As *HLA-DRB1* and *-DQB1* and *HLA-B* and *-C,* transmit together in northern Europeans [50], we used the *HLA-DQB1* typing results as a quality control to verify the accuracy of the *HLA-DRB1* allele callings, and to reduce the cost of genotyping; the *HLA-C* alleles were not detected in this study.

The existence of a recombination hotspot can justify the lack of LD between loci [2,272-274]: *HLA-A* gene lies physically far away from class II and a recombination spot is known to exist in the MHC class II region between *HLA-DRB1/DQB1* and *HLA-DPB1* [112,264]. Given that, only two extended and conserved *A~B~DRB1* haplotypes were observed; one with *HLA-DRB1*03:01* and another with *HLA-DRB1*13:02* [1,275]. *A*01-B*08-DRB1*03* haplotype is typically referred as the ancestral haplotype AH 8.1, or the autoimmune haplotype [1].

Table 7. *Significant (+) or non-significant (-) linkage disequilibrium (LD: r^2) between HLA genes*

HLA Locus	*A*	*B*	*DRB1*	*DQB1*
A		+	-	+
B	+		+	+
DRB1	-	+		+
DQB1	+	+	+	
DPB1	-	+	-	+

(Significance level=0.05) (Unpublished data)

5.1.4 THE EXTENDED MHC HAPLOTYPE ANALYSIS

Various SNPs within the MHC class III region have been described and implicated in disease association studies, as well as graft versus host disease, and variants in genes *TNF, C4, BTNL2* and *HLA-DRA* have been associated with disease predisposition [59,120,122]. However, only few studies have described the phase of different non-HLA variants together with the

traditional HLA alleles [91,276]. Here, we combined the SNPs and gene copy number data in MHC with those in the traditional HLA genes, assessed the extended MHC haplotypes, and finally investigated the LD and shared fragments between loci in relation to the polymorphic *HLA-DRB1*.

Fourteen of 74 genotyped SNPs were excluded due to low minor allele frequency (<0.01) or HWE p-value (<0.001). After quality control (QC), a total of 60 SNPs were accepted. Using a tag-SNP approach, we excluded additional five SNPs that were in total LD (r^2=1) with another genotyped SNP. The strong LD pattern indicates a high likelihood that there are other SNPs (not genotyped here) in strong LD with our genotyped SNPs. The marker haplotypes were constructed combining the SNPs of *TNF-LTA* (TNF block), *BTNL2* and *HLA-DRA* (BTNL2 block) and merging with the *C4A* and *C4B* gene copy numbers (C4 block). The rare *TNF, C4* and *BTNL2* haplotypes observed less than three times were excluded from the analysis. Taken together, five haplotypes covered 96% of TNF haplotypes, four haplotypes covered 99% of C4 haplotypes, and a further four haplotypes covered 80% of *BTNL2* haplotypes. Our results corroborated the previous publications anticipating that the number of frequent *TNF* and *BTNL2* haplotypes is limited, although the actual number of SNPs in the *TNF* and *BTNL2* region is high [276].

Table 8. *The strong and significant pairwise LD (D'/r^2)# between HLA-DRB1 alleles (observed n > 3) and MHC variants (traditional HLA alleles or TNF, C4 and BTNL2 blocks) (I)*

	TNF block	*C4 block*	*BTNL2 block*	*HLA-DQB1*	*HLA-DPB1*
DRB1*01:01	-	-	1.0/1.0 (BTNL2_1)	1.0/0.9 (*DQB1*05:01*)	-
DRB1*03:01	0.8/0.4 (TNF_3)	0.7/0.5 (C4_4)	1.0/1.0 (BTNL2_5)	1.0/0.7 (*DQB1*02*)	0.9/0.4 (*DPB1*01:01*)
DRB1*04:01	-	-	1.0/0.4 (BTNL2_3)	-	-
DRB1*08:01	-	-	1.0/0.8 (BTNL2_4)	1.0/1.0 (*DQB1*04*)	-
DRB1*09:01	-	-	-	1.0/0.8 (*DQB1*03:03*)	-
DRB1*11:01	-	-	0.9/0.8 (BTNL2_8)	-	-
DRB1*12:01	-	-	1.0/0.5 (BTNL2_10 and _11)	-	-
DRB1*13:01	-	-	1.0/0.6 (BTNL2_6)	1.0/1.0 (*DQB1*06:03*)	-
DRB1*13:02	0.9/0.5 (TNF_9)	0.9/0.6 (C4_5)	1.0/0.9 (BTNL2_7)	1.0/1.0 (*DQB1*06:04*)	-
DRB1*15:01	-	-	1.0/1.0 (BTNL2_2)	1.0/1.0 (*DQB1*06:02*)	-

Here the overall pairwise LD between loci is considered strong when D'> 0.8 (i.e., strong LD) and r^2 > 0.4 (moderate LD), together with a P value < 0.05.

The evaluation of the extended haplotypes containing *TNF, C4* and *BTNL2* blocks and *HLA-DRB1* suggested that the strongest pairwise LD

exists between *HLA-DRB1* and the *BTNL2* block (Table 9). Corresponding with the HLA haplotypes, only *HLA-DRB1*03:01* and *HLA-DRB1*13:02* formed conversed and extended MHC haplotypes covering also the *TNF* block (D'=0.8; D'=0.9, respectively) [1,50]. The LD and extended MHC haplotype analysis indicated that, in relation to *HLA-DRB1,* the highest degree of polymorphisms was observed in the MHC class I, TNF and C4 blocks (Table 8) and in the *HLA-DPB1* locus.

We observed that in a number of cases, the same variant (SNP or CNV) was present in many different extended *HLA-DRB1* haplotypes. This information can be used to further subgroup the *HLA-DRB1* alleles according to the presence of a common variant, e.g., in the promoter region of *LTA* or *TNF* or a splice-variant (e.g., A allele in rs2076530; Table 9). Furthermore, the *C4* deficiencies, previously associated with several inflammatory diseases [142,155,277-279], were observed within distinct extended *HLA-DRB1* haplotypes, C4A (C4_4) mainly with *HLA-DRB1*03:01,* and *C4B* (C4_2 and _3) with *HLA-DRB1*01:01, *04:01, *08:01* and **13:01*. Taking a closer look at the TNF and BTNL2 blocks, many of the SNPs or a SNP in strong LD (not genotyped here; http://www.snp-nexus.org) have been associated with several traits [280,281]. Furthermore, *HLA-DRB1*11:01* and a rare *HLA-DRB1*15:01*, both associated with inflammatory diseases, [214,282] shared same TNF haplotype (TNF_5).

Table 9. *A splice-site variant rs2076530 (A allele) in BTNL2 is present in different BTNL2 blocks and forms two-locus haplotypes (>1% presented here) with distinct HLA-DRB1 alleles*

BTNL2 SNP rs2076530 (A or G)	in different BTNL2 block	HLA-DRB1 allele (observed n >3)	Frequency of the DRB1-BTNL2 haplotype
G	*BTNL2_1*	*01:01	0.147
G	*BTNL2_3*	*04, *07	0.121
G	*BTNL2_6*	*13:01	0.053
A	*BTNL2_2*	*15:01	0.143
A	*BTNL2_4*	*08:01	0.103
A	*BTNL2_5*	*03:01	0.097
A	*BTNL2_7*	*13:02	0.040
A	*BTNL2_8*	*11:01	0.037
A	*BTNL2_10*	*12:01	0.020
A	*BTNL2_11*	*12:01	0.016
A	*BTNL2_12*	*07:01	0.013

5.2 MULTIPLE LEVELS OF MHC ASSOCIATION WITH SARCOIDOSIS

Previous linkage analyses and GWASs had shown a strong connection between MHC and sarcoidosis (Table 3 and 4). To further investigate the susceptibility loci, we performed a functional candidate gene analysis, focusing on MHC markers to discover variants that may contribute to the development of sarcoidosis and associate with the prognosis of sarcoidosis. We set out to replicate the known MHC associations with Finnish sarcoidosis patients and to locate novel susceptibility variants. To overcome the issue of sample heterogeneity, we subgrouped the sarcoidosis patients according to the progression of disease and Löfgren syndrome.

5.2.1 VARIANTS IN MHC CLASS II ASSOCIATE WITH SARCOIDOSIS (II, III)

Genes that encode the MHC class II molecules and are expressed on the surface of APC have been associated with sarcoidosis or its disease course (Table 3). We set out to confirm the association between *HLA-DRB1* and sarcoidosis in the Finnish sample, with the secondary aim of investigating the association between *HLA-DPB1* and sarcoidosis in general.

Exon 2, which encodes the peptide-binding region of *HLA-DRB1* and –*DPB1* was either sequenced or alleles were detected with the SSP method in 188 cases and 150 controls. All the genotypes and ambiguities were checked. We showed that variants of *HLA-DRB1* and –*DPB1* associated with sarcoidosis or its disease course (Table 10). We replicated the previously reported *HLA-DRB1*03:01* association with a favourable prognosis of sarcoidosis [224], showing that 34% of the patients with resolved disease had at least one copy of the allele, compared with 16% in the patients with persistent disease (P=0.007, OR=2.70, 95% CI=1.35–5.37). In addition, the extended MHC haplotype analysis presented a novel predisposing haplotype for sarcoidosis resolution, *HLA-DRB1*04:01-DPB1*04:01* (resolved 16.9% vs. persistent 6.1%; p=0.02, OR = 3.1, 95% CI =1.15–8.41) which was independent of other *HLA-DRB1* alleles. Our study showed that *HLA-DRB1*15:01* (sarcoidosis 41% vs. controls 28%; P= 0.016, OR= 1.76, 95 % CI=1.11-2.79) and *HLA-DPB1*04:01* (sarcoidosis 66% vs. controls 55%; p= 0.040, OR= 1.59, 95 % CI = 1.02-2.47) are general predisposing variants for sarcoidosis. The extended MHC haplotype analysis showed a predisposing effect for *DRB1*15:01-DPB1*04:01* haplotype (25 % vs. 16 %; P= 0.04, OR= 1.75, 95 %CI = 1.01-3.03). We could not successfully confirm the association of sarcoidosis with *HLA-DRB1*11, *12,* and the CBD-related marker of *HLA-DPB1*02:01* (also known as Glu69) as reported by others [211,215,219]. Two protective HLA phenotypes were found: we successfully replicated the association of *HLA-DRB1*01:01* (sarcoidosis 14% vs. controls 25%; P=0.008, OR=0.48, 95% CI = 0.27-0.83), and found a novel association with *HLA-*

*DPB1*04:02* (sarcoidosis 22% vs. controls 37%; P=0.003, OR = 0.48, 95% CI = 0.30–0.79).

Table 10. *Summary of significant associations between resolved and persistent sarcoidosis in the Finnish sample*

Variant	MHC class II molecule	Res (%)	Per (%)	P value	OR	CI 95%
DRB1*03:01 phenotype	*DRB1*03:01*	34	16	0.007	2.70	1.35-5.37
DRB1*04:01-DPB1*04:01 haplotype	*DRB1*04:01, DPB1*04:01*	16.7	6.1	0.02	3.07	1.13-8.29
Arginine at position 74	*DRB1*0301*	34	16	0.007	2.70	1.35-5.37
Tyrosine at position 26	*DRB1*03:01, *09:01*	39	24	0.027	2.10	1.10-3.40
Lysine at position 71	*DRB1*03:01, *04:01*	48	31	0.017	2.10	1.14-3.77
Asparagine at position 37	*DRB1*03:01, *09:01,*14:02,*13:02, *13:01*	62	42	0.006	2.30	1.28-4.11
Aspartic acid at position 70	*DRB1*07:01, *08:02, *08:01, *08:03, *11:01, *11:04, *12:01, *13:01, *13:02, *16:01*	52	72	0.006	0.42	0.23-0.76
Phenylalanine at position 67	*DRB1*08:01, *08:02, *09:01, *11:01, *11:04, *16:01*	29	45	0.025	0.50	0.27-0.91
Arginine at position 71	*DRB1*01:01, *01:02, *04:03, *04:04, *04:07, *04:08, *07:01, *08:01,*08:02, *08:03, *08:04, *09:01,*10:01, *11:01, *11:04,*12:01, *14:01, *14:02, *16:01*	57	79	0.002	0.36	0.19-0.65
Absent of DPB1*04:02 and presence of rs2075430 A allele	*DPB1*04:02*	65	79	0.041	0.05	0.27-0.98

Res, resolved sarcoidosis; Per,persistent sarcoidosis

To more extensively investigate the association of MHC class II molecules, we studied the relation of *HLA-DRB1* pocket residues and sarcoidosis. The amino acid sequence of selected *HLA-DRB1* alleles shows the polymorphism in more detail (Fig. 19). The analysis revealed that peptide binding sites 11, 13, 26, 28, 30, 47, 71 and 86, located in pockets 1, 4, 6 and 7, were associated with sarcoidosis. In contrast, peptide binding sites in positions 26, 37, 67, 70, 71 and 74, located in pockets 4, 7 and 9, were critical for disease prognosis (Table 11). Most interestingly, we showed that the arginine at position 71 was more common in patients with persistent disease, compared with the group with a resolved disease (see Table 11).

```
AA Pos.                  10         20         30         40         50         60         70         80         90         100
DRB1*01:01:01      GDTRPRFLWQ LKFECHFFNG TERVRLLERC IYNQEESVRF DSDVGEYRAV TELGRPDAEY WNSQKDLLEQ RRAAVDTYCR HNYGVGESFT VQRRVEPKVT
DRB1*03:01:01:01   --------EY STS------- -----Y-D-Y FH----N--- ------F--- ---------- ---------- K-GR--N--- -----V---- -----H----
DRB1*04:01:01      --------E- V-H------- -----F-D-Y F-H---Y--- ---------- ---------- ---------- K--------- ---------- -----Y-E--
DRB1*15:01:01:01   ---------- P-R------- -----F-D-Y F--------- ------F--- ---------- ------I--- A--------- -----V---- -----Q----

AA Pos.                 110        120        130        140        150        160        170        180        190        200
DRB1*01:01:01      VYPSKTQPLQ HHNLLVCSVS GFYPGSIEVR WFRNGQEEKA GVVSTGLIQN GDWTFQTLVM LETVPRSGEV YTCQVEHPSV TSPLTVEWRA RSESAQSKML
DRB1*03:01:01:01   ---------- ---------- ---------- ---------T --------H- ---------- ---------- ---------- ---------- ----------
DRB1*04:01:01      ---A------ ---------N ---------- ---------T ---------- ---------- ---------- ---------L ---------- ----------
DRB1*15:01:01:01   ---------- ---------- ---------- --L------- -M-------- ---------- ---------- ---------- ---------- ----------

AA Pos.                 210        220        230
DRB1*01:01:01      SGVGGFVLGL LFLGAGLFIY FRNQKGHSGL QPTGFLS
DRB1*03:01:01:01   ---------- ---------- ---------- --R----
DRB1*04:01:01      ---------- ---------- ---------- -------
DRB1*15:01:01:01   ---------- ---------- ---------- -------
```

Figure 19 The amino acid sequences of *HLA-DRB1*01:01, *03:01, *04:01* and **15:01* show a regions with variability. The sequences have been adopted from MGT/HLA Database. Sequence Alignment: Release 3.14.0 (2013-10-01) [78]

5.2.2 IS THE C4 ASSOCIATION REFLECTING UNDERLYING LD STRUCTURE? (II)

In addition to the traditional HLA genes, many promising non-HLA associations with different disease phenotypes and in different ethnic groups have been reported (summarized in Table 2). First, we studied the gene copy number of complement genes *C4A* and *C4B*, located within the MHC class III, as a novel susceptibility marker for sarcoidosis (Study II). The products of complement genes *C4A* and *C4B*, contribute in classical and lectin complement activation pathways, and the *C4* deficiencies have been previously associated with a number of autoimmune, inflammatory or infectious diseases [142,155,252,277,278,283,284]. The study set consisted of 188 Finnish sarcoidosis cases and 150 controls. The gene copy numbers of *C4A* and *C4B* and the CT insertion in *C4A* were detected with RT-PCR. Two subjects, one patient and one control, failed in the RT-PCR analysis, and were excluded.

The gene copy number of *C4A* and *C4B* and *C4-DRB1* haplotypes associated with sarcoidosis showed both predisposing (e.g., *C4A*1-C4B*1-DRB1*15:01*) and protective (e.g. *C4A*2-C4B*0-DRB1*01:01*) associations (Table 11). However, the logistic regression analysis indicated that *C4* associations were likely to be due to the haplotype structure with *HLA-DRB1*. Interestingly, with the extended MHC haplotype analysis we identified a novel, but rare, *HLA-DRB1*03:01* haplotype with one *C4A* gene that was associated with resolved disease (*C4A*1-C4B*1-DRB1*03:01*, $p = 0.031$, $OR = 7.89$, 95% CI = 0.96-64.77).

Table 11. *Complement C4 gene as a susceptibility marker for sarcoidosis (II)*

C4 number/deficiency/haplotype	Sarcoidosis (%)	Controls (%)	P value	OR	95% CI
C4A genes < 2 #	25	16	0.059	1.75	1.01-3.02
Haplotype: One C4A gene, one C4B gene, *DRB1*15:01*	20	11	0.004	1.92	1.24-2.97
C4A genes = 3	13	28	0.001	0.83	0.74-0.93
C4B genes < 2	29	41	0.021	0.58	0.37-0.91
No C4B gene	4	10	0.026	0.35	0.14-0.88
Haplotype: Two C4A genes, No C4B gene, *DRB1*01:01*	3	10	<0.001	0.29	0.14-0.57

#Including *C4A* insert if present

Typically, *HLA-DRB1*03:01* is found in an extended MHC haplotype with the C4A deficiency (AH.8.1). Surprisingly, only 66% of all *HLA-DRB1*03:01* alleles in the group with resolved sarcoidosis were inherited with the AH 8.1 haplotype [1], compared with 89% in persistent disease. It is therefore reasonable to speculate that the causal variant for sarcoidosis is in fact the *HLA-DRB1*03:01* allele (or another, unknown variant in LD with *HLA-DRB1*03:01*) and not the AH 8.1 haplotype.

5.2.3 REPLICATION OF *BTNL2* SPLICE-SITE VARIANT IN SARCOIDOSIS SUSCEPTIBILITY (III)

The association of *BTNL2,* a member of the immunoglobulin gene superfamily, with sarcoidosis has been studied in many populations with contradictory results [175,226,233,234,237]. The splice site variant A of rs2076530 located at the exon 5 could, in theory, result in abnormal T-cell regulation and antigen response [175]. In Study III with the Finnish sample (187 case and 150 controls; one resolved patient failed the SNP typing) we replicated the increase of carriership frequency of A (AA/AG) of rs2076530 [(rs2076530 (A)] with persistent sarcoidosis (92.9%) when compared with controls (84.0%) (AA/AG genotype; p=0.039, OR=2.48). In addition, the genotype AA and the allele frequency of A were significantly associated with the phenotype (unpublished results: carrier of A allele: p=0.01, OR=1.62; AA genotype; p=0.048, OR=1.68).

The strong LD between *HLA-DRB1* and rs2076530, as observed in our study (Table 12), complicated the detection of the causal effect. Indeed, *HLA-DRB1*01* is in total LD (D'=1.0) with rs2076530 (G) and *HLA-DRB1*03* and **15* in strong LD (D'>0.8) with rs2076530 (A). We were able to show an independent role for both *BTNL2* rs2076530 (A) as a predisposing and MHC class II as a protecting (*HLA-DRB1*01:01* and *DPB1*04:02*) marker in an extended MHC haplotype analysis, where *HLA-DRB1*, *HLA-DPB1* and rs2076530 were combined, and in a logistic regression analysis.

Table 12. *The LD between BTNL2 rs2076530 MHC class II alleles (HLA-DRB1 and –DPB1) in sarcoidosis patients (all, resolved, persistent) and controls*

LD between DRB1/DPB1 and rs2075430	Sarcoidosis		Resolved		Persistent		Controls	
rs2076530	A	G	A	G	A	G	A	G
DPB1*01:01	0.7		0.8				0.9	
DPB1*02:01								0.4
DPB1*03:01	0.4						0.8	
DPB1*04:01								
DPB1*04:02								0.2
DRB1*01:01		1.0		1.0		1.0		1.0
DRB1*03:01	0.9		1.0		0.8		1.0	
DRB1*04:01		0.8		0.9		0.7		0.9
DRB1*08:01	0.8		0.5		0.8		0.9	
DRB1*15:01	1.0		1.0		1.0		0.9	

D'>0.8, Strong LD; >0.4 D' < 0.8, Moderate LD; >0.2 D'< 0.4, Low LD

5.2.4 COMMON AND POPULATION-SPECIFIC MHC VARIANTS ASSOCIATE WITH SARCOIDOSIS (IV)

The aim of the Sarcoidosis Collaboration Study was to replicate the association of MHC SNP variants to sarcoidosisin Finnish sarcoidosis patients and to explore associations shared across different populations or population-specific features. We used the functional candidate gene approach to identify genetic susceptibility loci for sarcoidosis. The genes selected (*AGER, LTA, TNF, BTNL2, HLA-DRA*) were finemapped with tag-SNPs covering the genes from 5' to 3' region. The previously replicated SNP rs2076530 was also included in Study IV.

After assessing the association between SNPs and sarcoidosis in the discovery sample (188 cases and 150 controls) we performed a replication and meta-analysis in four European populations (a total of 805 sarcoidosis patients and 870 controls, including the Finnish discovery sample). To reduce the study heterogeneity, patients were subgrouped to Löfgren (LS) and non-Löfgren patients (NL), and further to non-Löfgren resolved (NLR) and non-Löfgren persistent sarcoidosis (NLP) based on the disease outcome. The quality control procedure is detailed in Materials and Methods.

Sixteen SNP were selected for replication, based on the analysis of the Finnish discovery sample. In the sample, 13 SNPs showed evidence of association ($P < 0.05$) with either NL (seven SNPs; Fig. 20), LS (two SNPs) or the disease course of sarcoidosis (six SNPs). Three additional SNPs (rs3130349, rs1800624 and rs3135365) were selected based on nominal evidence of association. Furthermore, in the discovery sample, the strongest

evidence of association was observed with NL and rs3177928 (*HLA-DRA*; p=0.001, OR=2.17), and NL and rs28362677 (*BTNL2*; p=0.002, OR=1.92).

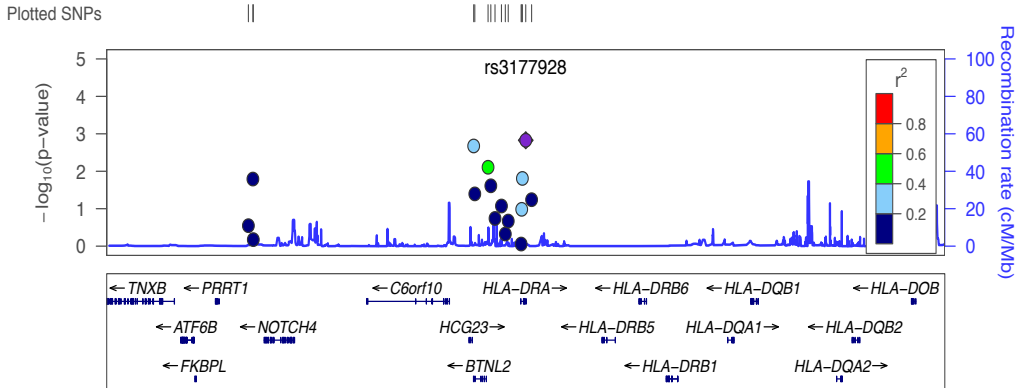


Figure 20 Seven SNPs showed association (uncorrected P < 0.05) with non-Löfgren (NL) sarcoidosis in the Finnish discovery sample, the most significant being rs3177928 (shown as a purple square in the fig.). Other SNPs are shown with dots; the color of the dots represents the extent of r^2 with the most significant SNP, rs3177928. Data from the 1000 Genomes project was used for recombination rates and LD pattern [21].

In the meta-analysis, several markers showed evidence of association with low heterogeneity (I^2<25) (Table 13). In addition, population-specific MHC gene variants associating with the disease phenotype were found. The strongest association was for NL sarcoidosis and was detected with rs3177928 (P = 1.79E-07, OR=1.89, I^2=0) near the *HLA-DRA* gene.

In patients with LS or NL patients with resolved disease, the strongest association was observed in *HLA-DRA* gene with SNP rs3129843 (P= 3.44E-12, OR=3.44, I^2=11.54; P= 0.0029, OR=1.86, I^2=12.78, respectively).

Since the MHC region is known for its strong LD pattern, we investigated whether the MHC associated markers were independent of the association within the MHC region. We performed a conditional stepwise association analysis among the variants with $P < 0.05$ in the meta-analysis, using the HLA-DRB1 alleles as the covariate. Four of the observed variants near genes HLA-DRA and BTNL2 were independent of the HLA-DRB1 alleles, and were associated with distinct disease phenotypes with significant residual associations (bolded in Table 14). Furthermore, significant predisposing and independent variants for NL (vs. controls) were rs3135365, rs3177928, HLA-DRB1*15 and the absence of HLA-DRB1*16; for LS (vs. controls), these were rs6937545 and HLA-DRB1*03, *13 and *14; and for NLR (vs. NLP), they were rs5007259 and HLA-DRB1*01. In addition, rs3135351 was independent of HLA-DRB1 alleles but not independent of rs5007259 (in analysis of NLR vs. NLR). However, some associations observed in the meta-analysis, like the association with rs3129843 and LS, were secondary to the HLA-DRB1*03:01 allele (i.e., the SNP and HLA-DRB1*03:01 are in strong LD and the

association of HLA-DRB1*03:01 with the trait was more significant than with the rs3129843 SNP).

Table 13. *The meta-analysis of subjects in the Sarcoidosis Collaboration Study showed that variants in AGER, HLA-DRA and BTNL2 are associated with the disease phenotype ($p<0.05$, $I^2<25$). When adjusting for HLA-DRB1, SNPs in the non-coding region of HLA-DRA and BTNL2 remained significant (**bolded**)*

NL (n=629) vs. controls (n=859)							
SNP	Gene(s)	Function	A1	A2	P(R) value	OR	I^2
rs1800684	AGER	Coding exon	A	T	0.0034	0.73	0
rs2076530	BTNL2	Coding exon	A	G	0.0003	1.34	0
rs3763313	BTNL2	Promoter	A	C	1.24E-06	1.63	0
rs9268528	BTNL2	Promoter	A	G	0.2450	0.91	0
rs3135365	**BTNL2/DRA**		**T**	**G**	**0.009398**	**0.78**	**0**
rs3129877	HLA-DRA	Intron	A	G	0.0197	0.82	3.3
rs3135392	HLA-DRA	Intron	T	G	0.0058	0.81	0
rs3177928	**HLA-DRA**	**Downstream**	**G**	**A**	**1.79E-07**	**1.90**	**0**
LS (n=136) vs. controls (n=859)							
SNP	Gene(s)	Function	A1	A2	P(R) value	OR(R)	I^2
rs3130349	RNF5, AGPAT1, AGER	Coding exon, 5upstream, 3' UTR	G	A	2.03E-10	0.39	0
rs1800624	AGER, PBX2	Promoter, 3' UTR	T	A	0.0013	1.70	0
rs2076530	BTNL2	Coding exon	A	G	0.0086	1.45	0
rs3763313	BTNL2	Promoter	A	C	0.0003	2.05	0
rs5007259	BTNL2	Promoter	T	C	7.63E-07	1.99	0
rs3129843	BTNL2/DRA		G	A	3.44E-12	3.44	11.5
rs9268644	HLA-DRA	Intron	A	C	0.0010	1.69	21.3
rs6937545	**HLA-DRA**	**Downstream**	**A**	**C**	**1.00E-09**	**2.28**	**0**
NLP (n=249) vs. NLP (n=337)							
SNP	Gene(s)	Function	A1	A2	P(R) value	OR(R)	I^2
rs28362677	BTNL2	Coding exon	G	A	0.0047	0.57	0
rs5007259	**BTNL2**	**Promoter**	**T**	**C**	**0.0412**	**1.28**	**0**
rs3135351	BTNL2/DRA		T	G	0.0038	1.53	0
rs3129843	BTNL2/DRA		G	A	0.0030	1.86	12.8
rs3129877	HLA-DRA	Intron	A	G	0.0048	1.45	0

To explore the extended MHC haplotypes more thoroughly, we created extended MHC haplotypes with *HLA-DRB1* alleles and the 16 genotyped SNPs (Table 14 and 15). All the subjects that participated in the Sarcoidosis Collaboration Study and had the *HLA-DRB1* genotypes available were included into the analysis (sarcoidosis patients n=461, controls=516). Subjects with missing SNP genotypes were excluded from the analysis. We

did not have the *HLA-DRB1* status available for the Dutch sample set. The evidence favours the explanation that there is a wide range of polymorphism in the haplotypes. Furthermore, some of the variation is clearly due to population-specific features (Table 14). In Table 15, the full *16-SNP-DRB1* haplotypes are presented, showing that polymorphisms in MHC class II-III region with particular *HLA-DRB1* alleles are observed. It is reasonable to speculate that both the non-HLA genes (and *HLA-DRA*) and the antigen-presenting molecule (*HLA-DRB1)* are needed for the susceptibility of sarcoidosis. Here, we did not perform a subphenotype analysis, due to sample size limitations.

Table 14. *The haplotype frequency (2n) of the 16-SNP-DRB1 haplotypes in Finnish (Fin), Swedish (Swe) and Czech samples (Cze) shows population specific differences both in sarcoidosis patients (NL+LS) and controls*

SNP haplotype	with HLA-DRB1	NL+LS Fin n=187	Controls Fin n=150	NL+LS Swe n=184	Controls Swe n=187	NL+LS Cze n=90	Controls Cze n=179
1	*01	0.06	0.14	0.03	0.08	0.03	0.05
2	*01	0.01	0.01	0.01	0.02	0.02	0.02
3	*03	0.12	0.09	0.19	0.12	0.13	0.09
4	*04	0.06	0.08	0.08	0.10	0.03	0.03
5	*04	0.05	0.03	0.06	0.05	0.05	0.06
6	*07	0.03	0.02	0.01	0.01	0.03	0.04
7	*07	0.01	0.01	0.01	0.04	0.01	0.03
8	*07	-	0.01	0.02	0.04	0.01	0.04
9	*08	0.08	0.11	0.04	0.02	0.02	0.02
10	*08	0.01	-	0.01	0.01	0.01	-
11	*09	0.01	0.01	-	0.01	-	-
12	*09	0.01	0.02	-	-	-	-
13	*11	0.06	0.04	0.05	0.03	0.06	0.08
14	*11	-	-	0.01	0.01	0.02	0.03
15	*12	0.01	0.02	0.02	0.02	0.02	0.02
16	*12	0.01	0.02	-	-	-	-
17	*13	0.03	0.05	0.03	0.02	0.04	0.01
18	*13	0.03	0.03	0.02	0.04	0.02	0.01
19	*13	0.03	0.02	0.01	0.01	0.04	0.03
20	*13	0.01	0.01	0.02	0.01	0.03	0.03
21	*13	-	-	0.01	-	0.02	0.02
22	*13	0.01	0.01	-	-	-	-
23	*14	0.01	-	0.02	0.02	0.02	0.02
24	*14	0.01	-	0.02	0.01	0.03	-
25	*15	0.17	0.11	0.10	0.10	0.13	0.10
26	*15	0.03	0.03	0.03	0.03	0.03	0.01
27	*15	-	-	0.01	-	0.01	-
28	*16	-	0.01	-	0.01	0.04	0.03

NL, Non-Löfgren; LS, Löfgren syndrome

Table 15. *Extended sixteen-SNP and HLA-DRB1 haplotypes in sarcoidosis patients (n= 461) and controls (n=516)#*

SNP	Gene (Function)	HLA-DRB1 allele																
		*01	*01	*03	*04	*04	*07	*07	*07	*08	*08	*09	*09	*11	*11	*12	*12	*13
rs3130349	AGER (3' UTR)	G	G	A	G	G	A	A	G	G	G	G	G	G	G	G	G	G
rs1800684	AGER (Coding exon)	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
rs1800624	AGER (Promoter)	T	A	T	T	A	T	T	T	A	T	A	T	A	T	A	A	T
rs2836677	BTNL2 (Coding exon)	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A
rs2076530	BTNL2 (Coding exon)	G	G	A	G	G	G	A	G	A	A	A	A	A	A	A	A	G
rs3763313	BTNL2 (Promoter)	C	C	A	A	A	A	C	A	C	C	A	A	A	A	A	A	A
rs5007259	BTNL2 (Promoter)	C	C	T	C	C	C	C	C	C	C	T	C	T	T	T	T	C
rs9268528	BTNL2 (Promoter)	A	A	A	G	G	G	A	A	A	A	G	G	G	G	G	G	A
rs3135365	BTNL2/DRA	T	T	T	T	T	T	G	T	T	T	T	T	T	T	T	T	T
rs3135351	BTNL2/DRA	G	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	T
rs3129843	BTNL2/DRA	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A
rs9268644	HLA-DRA (Intron)	A	A	A	C	C	C	C	C	C	C	C	C	C	C	C	A	C
rs3129877	HLA-DRA (Intron)	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs3135392	HLA-DRA (Intron)	T	T	T	G	G	G	T	G	T	T	G	G	G	G	T	T	G
rs3177928	HLA-DRA (3' UTR)	A	A	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G
rs6937545	HLA-DRA (3' UTR)	C	C	A	C	C	C	C	C	C	C	C	C	C	C	C	C	A
Haplotype frequency	Sarcoidosis	4.2%	1.2%	14.5 %	5.8%	5.0%	2.3%	1.1%	1.0%	4.5%	0.7%	0.5%	0.4%	0.9%	1.5%	0.4 %	3.3%	2.4%
	Controls	9.0%	1.8%	10.3 %	7.2%	4.6%	2.5%	2.6%	3.0%	4.9%	0.3%	0.5%	0.6%	5.3%	1.3%	1.9%	0.7 %	2.5%

SNP	Gene (Function)	HLA-DRB1 allele																
		*13	*13	*13	*13	*13	*13	*14	*14	*15	*15	*15	*15	*15	*15	*16	*16	*16
rs3130349	AGER (3' UTR)	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs1800684	AGER (Coding exon)	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A
rs1800624	AGER (Promoter)	T	T	A	T	A	A	A	T	T	T	T	T	T	T	A	A	A
rs2836677	BTNL2 (Coding exon)	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs2076530	BTNL2 (Coding exon)	G	G	G	G	A	G	A	A	A	A	A	A	A	A	A	A	A
rs3763313	BTNL2 (Promoter)	C	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
rs5007259	BTNL2 (Promoter)	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T
rs9268528	BTNL2 (Promoter)	A	A	A	G	G	G	G	G	A	A	A	A	A	A	A	A	A
rs3135365	BTNL2/DRA	T	T	T	T	T	T	T	T	G	G	G	G	G	G	T	T	T
rs3135351	BTNL2/DRA	G	G	T	T	G	G	G	G	G	G	G	G	G	G	G	G	G
rs3129843	BTNL2/DRA	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
rs9268644	HLA-DRA (Intron)	A	A	A	C	C	C	C	C	A	A	A	A	A	A	C	C	C
rs3129877	HLA-DRA (Intron)	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs3135392	HLA-DRA (Intron)	T	T	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs3177928	HLA-DRA (3' UTR)	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs6937545	HLA-DRA (3' UTR)	C	C	A	C	C	C	C	C	A	A	A	A	A	A	A	A	A
Haplotype frequency	Sarcoidosis	2.2%	0.9%	0.5%	1.6%	1.7%	13.4%	3.1%	0.7%	1.3%								
	Controls	1.9%	0.6%	0.4%	1.1%	0.3%	10.1%	2.3%	0.0%	1.4%								

#The subjects that participated in the Sarcoidosis Collaboration Study and had the *HLA-DRB1* data available

In contrast to Study III, the *BTNL2* (rs2076530) association did not reach the level of significance in the meta-analysis, after adjusting for the *HLA-DRB1* alleles. Furthermore, the SNP markers associated with disease phenotypes and showing heterogeneity (I^2>25) indicated that population-specific associations may occur. Table 16 presents the results of a case-control association analysis within each population, using three different disease phenotypes (NL, LS, NLR and NLP).

Table 16. *Several SNPs showing significant (or nominal) evidence of association in the discovery sample (in three different analyses) were replicated in at least one of the European studies participating in the Sarcoidosis Collaboration Study. In addition, population-specific association were observed.*

Non-Löfgren (NL) sarcoidosis vs. Controls

NL vs. Control	**Finnish (168 vs.150)**		**Swedish (112 vs. 358)**		**Dutch (180 vs. 180)**		**Czech (169 vs.178)**	
SNP(gene)	P	OR	P	OR	P	OR	P	OR
rs3130349 (AGER)	0.284	0.79	0.1016	1.44	0.5873	1.11	0.7372	1.07
rs1800684 (AGER)	0.016	0.60	0.1145	0.71	0.6758	0.91	0.1328	0.71
rs1800624 (AGER)	0.670	0.93	0.07818	0.75	0.9788	1.00	0.9455	1.01
rs28362677 (BTNL2)	0.002	1.92	0.1849	1.37	0.0001563	2.48	0.4551	1.18
rs2076530 (BTNL2)	0.040	1.41	0.05729	1.35	0.2079	1.22	0.03493	1.40
rs3763313 (BTNL2)	0.008	1.63	0.02956	1.57	0.02766	1.54	0.006012	1.84
rs5007259 (BTNL2)	0.024	1.43	0.01265	1.47	0.8397	0.97	0.000805	1.68
rs9268528 (BTNL2)	0.182	0.80	0.9733	0.99	0.2507	0.84	0.9078	1.02
rs3135365 (BTNL2/DRA)	0.085	0.70	0.06114	0.71	0.3698	0.84	0.4726	0.87
rs3135351 (BTNL2/DRA)	0.474	1.15	0.5828	0.90	0.5807	0.90	0.08462	1.42
rs3129843 (BTNL2/DRA)	0.214	1.37	0.2343	0.74	0.4393	0.84	0.7248	1.10
rs9268644 (DRA)	0.878	0.98	0.8369	1.03	0.1	0.78	0.07208	1.32
rs3129877 (DRA)	0.104	0.76	0.1016	0.76	0.07022	0.75	0.6903	1.07
rs3135392 (DRA)	0.016	0.68	0.149	0.80	0.1105	0.78	0.9523	0.99
rs3177928 (DRA)	0.001	2.17	0.004588	2.10	0.01985	1.64	0.01984	1.85
rs6937545 (DRA)	0.057	1.36	0.08723	1.31	0.422	0.88	0.01959	1.44

NL,Non-Löfgren; LS,Löfgren; NLR,non-Löfgren syndrome and resolved disease; NLP,non-Löfgren and persistent disease

Löfgren (LS) sarcoidosis vs. Controls

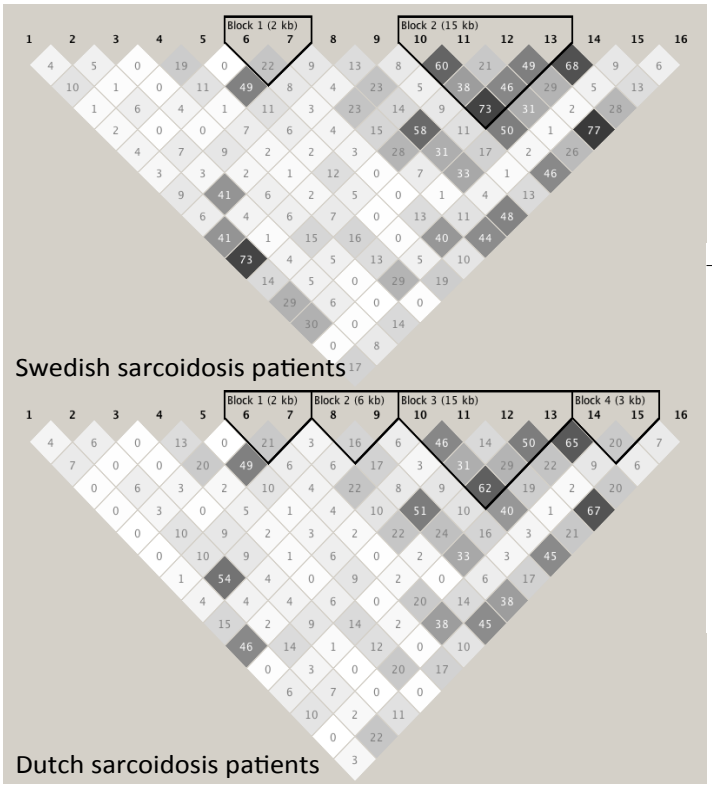
LS vs. Controls	**Finnish (19 vs. 150)**		**Swedish (78 vs. 358)**		**Dutch (- vs. -)**		**Czech (39 vs. 178)**	
SNP	P	OR	P	OR	P	OR	P	OR
rs3130349 (AGER)	0.199	0.58	2.17E-08	0.35	-	-	0.001061	0.42
rs1800684 (AGER)	0.199	0.58	0.3968	1.28	-	-	0.04788	3.17
rs1800624 (AGER)	0.279	1.54	0.0248	1.65	-	-	0.03247	1.87
rs28362677 (BTNL2)	0.6135	1.25	0.06064	1.74	-	-	0.1497	0.63
rs2076530 (BTNL2)	0.2144	1.59	0.01271	1.59	-	-	0.5734	1.16
rs3763313 (BTNL2)	0.1475	1.87	0.001464	2.38	-	-	0.1685	1.68
rs5007259 (BTNL2)	0.03544	2.14	0.0001564	1.99	-	-	0.01157	1.94
rs9268528 (BTNL2)	0.4936	1.32	0.5801	1.11	-	-	0.002022	2.40
rs3135365 (BTNL2/DRA)	0.05521	0.48	0.06476	1.64	-	-	0.0114	3.20
rs3135351 (BTNL2/DRA)	0.5483	1.27	8.35E-09	2.86	-	-	6.03E-09	4.52
rs3129843 (BTNL2/DRA)	0.2432	1.75	7.74E-12	3.78	-	-	5.29E-06	3.85
rs9268644 (DRA)	0.2353	1.53	0.04055	1.44	-	-	0.0006642	2.34
rs3129877 (DRA)	0.6495	0.85	0.0008453	1.81	-	-	1.10E-06	3.38
rs3135392 (DRA)	0.295	0.70	0.09121	1.35	-	-	1.26-05	2.99
rs3177928 (DRA)	0.1489	2.39	0.0006253	3.39	-	-	0.841	1.08
rs6937545 (DRA)	0.00988	2.43	6.30E-05	2.03	-	-	3.20E-05	2.83
rs3177928 (DRA)	0.8042	1.10	0.7893	1.16	0.6259	1.17	0.04049	0.38
rs6937545 (DRA)	0.004038	1.89	0.9734	1.01	0.4446	1.18	0.8744	0.96

non-Löfgren syndrome and resolved disease (NLR) vs. non-Löfgren and persistent disease (NLP)

NLR vs. NLP	Finnish (79 vs. 89)		Swedish (33 vs. 75)		Dutch (90 vs. 90)		Czech (47 vs. 83)	
SNP	P	OR	P	OR	P	OR	P	OR
rs3130349 (AGER)	0.3541	0.76	0.5253	0.76	0.2175	0.71	0.4221	1.30
rs1800684 (AGER)	0.2179	0.72	0.9727	1.01	0.6603	1.14	0.5487	0.80
rs1800624 (AGER)	0.07906	1.51	0.1758	0.66	0.7198	1.09	0.189	1.44
rs28362677 (BTNL2)	0.05861	0.54	0.8676	0.93	0.02773	0.39	0.1777	0.60
rs2076530 (BTNL2)	0.2253	0.75	0.3836	1.31	0.5785	1.13	0.7644	1.09
rs3763313 (BTNL2)	0.3511	1.29	0.6565	1.21	0.2377	0.70	0.4345	0.73
rs5007259 (BTNL2)	0.2278	1.31	0.1877	1.49	0.2041	1.31	0.7914	1.07
rs9268528 (BTNL2)	0.01434	1.78	0.4229	0.79	0.5176	1.15	0.2416	1.36
rs3135365 (BTNL2/DRA)	0.2155	0.72	0.4867	1.29	0.2381	1.36	0.2187	0.68
rs3135351 (BTNL2/DRA)	0.01629	1.85	0.2908	1.47	0.1212	1.50	0.5839	1.20
rs3129843 (BTNL2/DRA)	0.00407	2.63	0.141	2.00	0.04417	2.00	0.9591	0.98
rs9268644 (DRA)	0.004265	1.89	0.9672	1.01	0.5969	1.12	0.1256	1.49
rs3129877 (DRA)	0.02308	1.71	0.6702	1.15	0.1675	1.38	0.1798	1.48
rs3135392 (DRA)	0.2571	1.28	0.7688	0.91	0.2361	1.29	0.332	1.30
rs3177928 (DRA)	0.8042	1.10	0.7893	1.16	0.6259	1.17	0.04049	0.38
rs6937545 (DRA)	0.004038	1.89	0.9734	1.01	0.4446	1.18	0.8744	0.96

Many of the SNPs that showed evidence of association with the trait in the Finnish discovery sample were replicated in the independent European samples. One can argue that the sample groups are rather small after subgrouping the patients according to disease phenotype. Hence, the association between the variant and trait did not always reach the level of significance, probably due to the limited sample size, highlighting the importance of using meta-analysis to obtain larger subgroups and more statistical power for future studies. We also discovered clear population-specific differences between the analyses. Most importantly, the exonic BTNL2 SNP rs28362677 was replicated in the Dutch sample, showing strong evidence of associating with both sarcoidosis and with the disease course (P = 0.00016, OR=2.48; P = 0.028, OR= 0.39, respectively), which was not seen in the Swedish or Dutch sample.

The SNP LD blocks analysis of Finnish, Swedish, Dutch and Czech samples confirms the population-specific LD patterns, indicating that the Swedish have more conserved and longer LD patterns in MHC class II-III region than, for example, the Dutch (Fig. 21).



Marker	Candidate genes
1 rs3130349	RNF5, AGPAT1, AGER
2 rs1800684	AGER
3 rs1800624	AGER, PBX2
4 rs28362677	BTNL2
5 rs2076530	BTNL2
6 rs3763313	BTNL2
7 rs5007259	BTNL2
8 rs9268528	BTNL2
9 rs3135365	BTNL2/DRA
10 rs3135351	BTNL2/DRA
11 rs3129843	BTNL2/DRA
12 rs9268644	HLA-DRA
13 rs3129877	HLA-DRA
14 rs3135392	HLA-DRA
15 rs3177928	HLA-DRA
16 rs6937545	HLA-DRA

Figure 21 The extended SNP block (r^2) in Swedish sarcoidosis patients (NL+LS) is longer than in the Dutch sarcoidosis patients (NL+LS), suggesting population-specific features in the MHC profiles and distinct recombination sites, where the LD breaks down.

6 DISCUSSION

6.1 MHC PROFILE OF THE FINNS

Extended MHC haplotyping is a valuable tool for detecting disease association as well as anthropological studies. It provides a deeper understanding of the associations found in the MHC region and assists in characterizing the differences in LD structure between different populations [50,86,272,275,285]. The results of Study I provided new information about the correlation between traditional HLA alleles and MHC variants in the Finnish sample. The Finns' extended MHC haplotypes, covering both traditional HLA genes and non-HLA genes, have not been investigated before in depth. Although no novel alleles were detected in this study, the Finnish allelic and common haplotypic MHC profile was characterized.

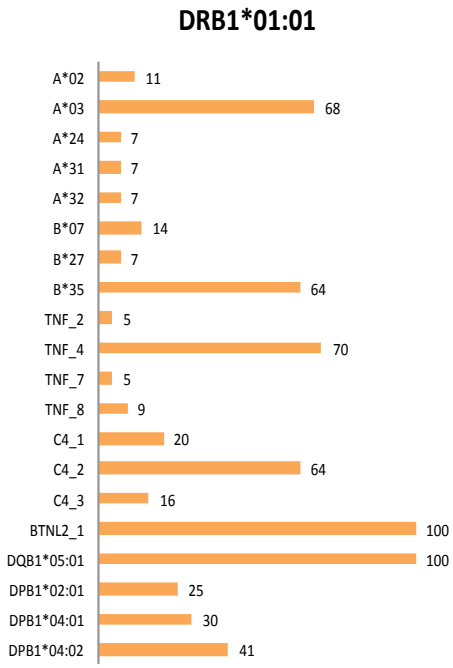
The enrichment of *A*03-B*35-DRB1*01* haplotype and the infrequency of the common European haplotype, *A*01-B*08-DRB1*03*, (AH 8.1) represents the unique genetic heritage of the Finns [112,286,287]**.** The 16th International HLA and Immunogenic Workshop IHIW project "Analysis of HLA Population Data" [112], in which we collaborated, performed the HLA allele frequency comparisons between populations. The results showed that the Finns and the Sami were genetically closer to the North-East Asians than to other European populations corresponding to the previous non-HLA anthropological analyses [112,287]. Overall, the similarities between the Finns and the Sami indicated that the distribution of HLA alleles follows the north-to-southeast axis, which is comparable with the previous Finnish HLA studies, which show large regional variations between different geographical areas of Finland. These regional differences may be caused by natural barriers such as lakes or a particular immune response to a particular pathogen [288,289]. Whereas the previous publication of the regional HLA diversity in Finland was mainly based on serological typing [288], a more comprehensive study of the MHC regional variation (of both HLA alleles and extended haplotypes) should be conducted [288,290].

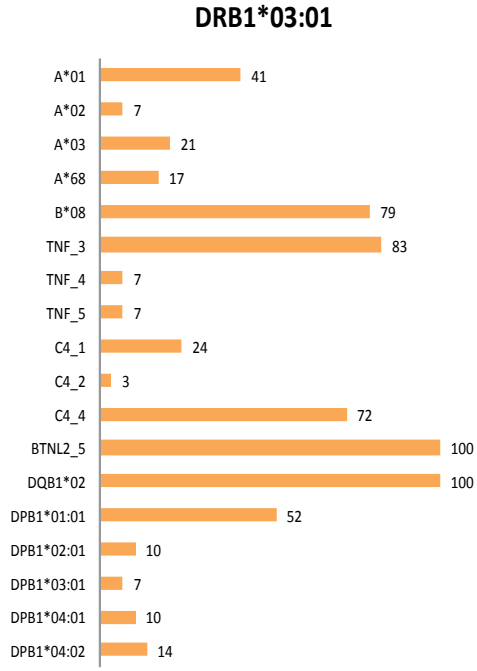
The main result of this study was the Finnish MHC profile, together with the two-locus haplotype map presented in Fig. 21, which characterizes the diversity of the Finnish MHC haplotypes in relation to *HLA-DRB1*. The MHC profile, togeth with the haplotypes of *HLA-DRB1,* can be used as guidance to explore the common allelic diversity of the neighbouring genes and the LD pattern of the region. Moreover, population specific haplotypes can lead to a better understanding of disease risk [224,255]. The widespread variation observed in the MHC profile provides essential information of the LD structure, which assists the interpretation of association signals in human traits and suggests future studies. For example, if the most significant SNP is localized at the '3 end of *HLA-DRB1* gene, and *HLA-DRB1*15* is previously

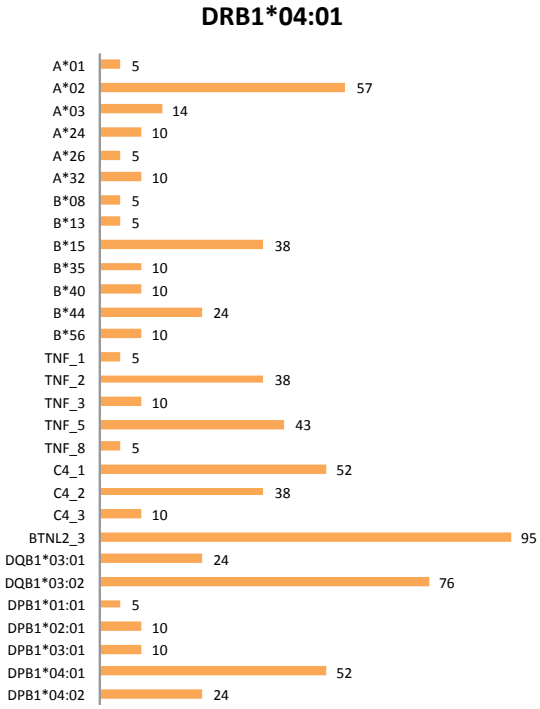
suggested as a disease susceptibility variant, one could recommend not to further analyse BTNL2 or HLA-DQB1 loci that are almost always in tight LD with HLA-DRB1*15. This makes detectingthe independent effect a challenging task, but instead analysing HLA-DPB1, which shows wider allelic diversity and multiple different haplotypes with HLA-DRB1*15 allele. Moreover, the characterization of the HLA-DRB1*03 haplotypes in the Finns agreed with the previous observation that most of the polymorphisms of the DR-haplotype are observed in either the 5' end (near HLA-A) or the 3' region (near HLA-DPB1). The need to adjust for HLA alleles and to analyse residual associations can also be estimated with the knowledge of population-specific MHC profiles. The VITA sample, representing the Finnish population and used in this study, was collected in Helsinki, Finland, and the distribution of the traditional HLA alleles was consistent with the previously published reports including thousands of Finnish individuals from a registry for bone-marrow transplantations (BMTs) [288,290,291]. This suggests that a population's common HLA allele frequencies can be estimated with modest sample collected from metropolitan area, such as the one used in this study. However, as mentioned before, regional differences exist, and warrant more studies.

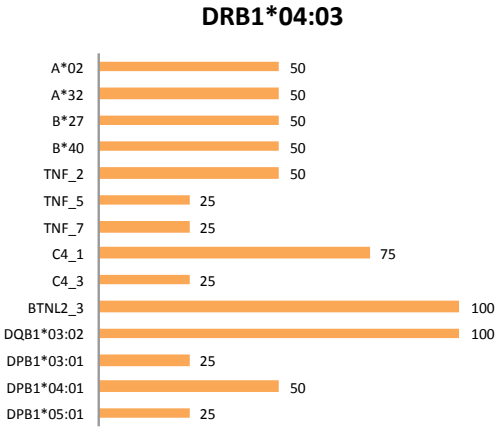
The population-specific haplotypes are also advantageous for matching patients for transplantation and may reveal regional restrictions for the HLA alleles [264,292]. As an example, Eastern European Americans and Eastern Europeans share similarities in the *HLA-DRB1* frequency, but the multi-locus haplotypes display regional differences between them [264,293]. In addition, understanding the full HLA profile of a population can lead to better stratification of individuals (e.g., for control selection) that share common genetic heritage. The knowledge of a population's MHC profile distribution may also improve the HLA matching in transplantation. With the ancestry information, the donors HLA haplotypes can be predicted from low-resolution types [120,294].

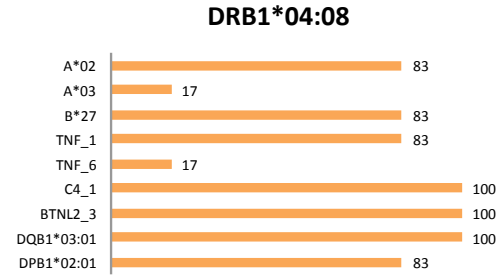
Multi-population collaborations to collect HLA-typing information from different populations and subpopulations allow illustrating detailed MHC profiles and genetic maps of the human population, as well as estimating regional differences within the population. To date, Europe is well characterized for HLA data, while only a few HLA-typed population samples are available from Africa. The HLA genetic landscape of Africa offers a huge potential to detect widespread variation in the MHC and may provide novel insights into the human past [112,289,293,295,296].

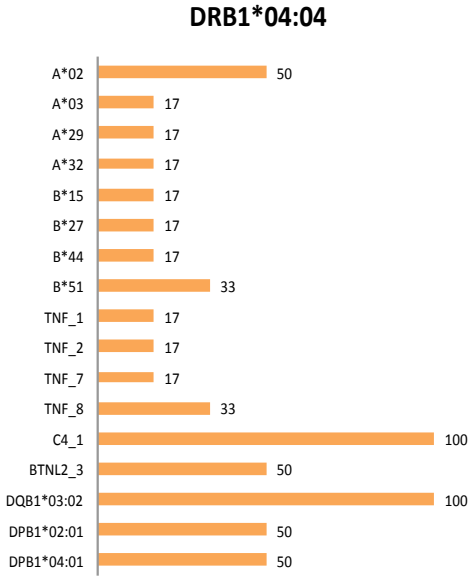

DRB1*01:01
A*02 11
A*03 68
A*24 7
A*31 7
A*32 7
B*07 14
B*27 7
B*35 64
TNF_2 5
TNF_4 70
TNF_7 5
TNF_8 9
C4_1 20
C4_2 64
C4_3 16
BTNL2_1 100
DQB1*05:01 100
DPB1*02:01 25
DPB1*04:01 30
DPB1*04:02 41

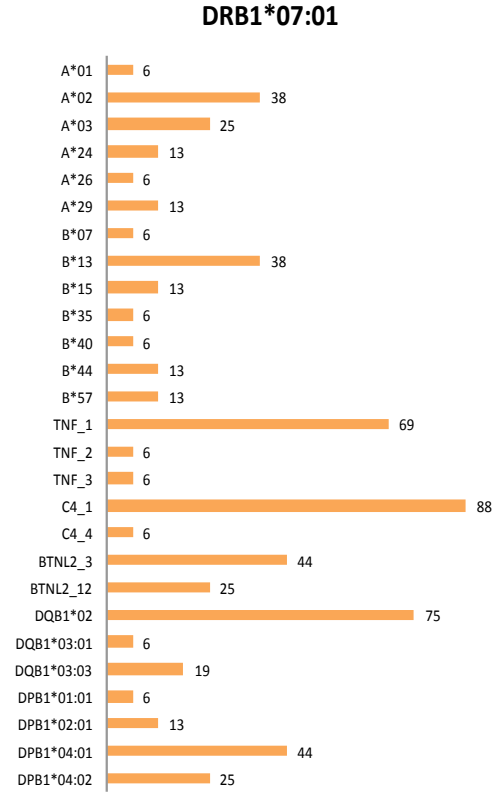

DRB1*03:01
A*01 41
A*02 7
A*03 21
A*68 17
B*08 79
TNF_3 83
TNF_4 7
TNF_5 7
C4_1 24
C4_2 3
C4_4 72
BTNL2_5 100
DQB1*02 100
DPB1*01:01 52
DPB1*02:01 10
DPB1*03:01 7
DPB1*04:01 10
DPB1*04:02 14

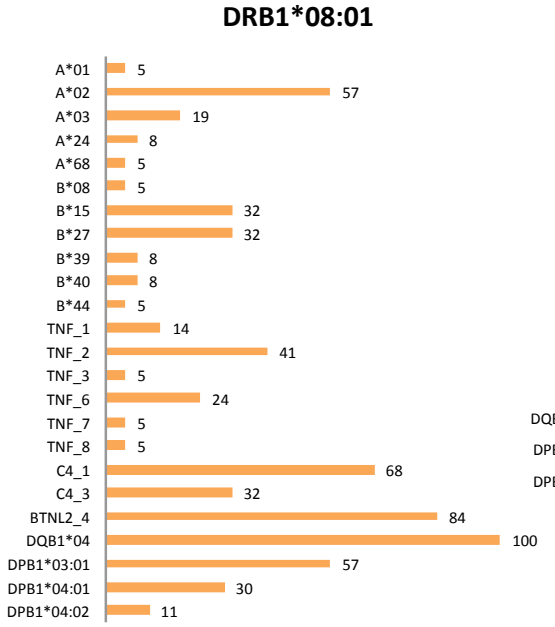

DRB1*04:01
A*01 5
A*02 57
A*03 14
A*24 10
A*26 5
A*32 10
B*08 5
B*13 5
B*15 38
B*35 10
B*40 10
B*44 24
B*56 10
TNF_1 5
TNF_2 38
TNF_3 10
TNF_5 43
TNF_8 5
C4_1 52
C4_2 38
C4_3 10
BTNL2_3 95
DQB1*03:01 24
DQB1*03:02 76
DPB1*01:01 5
DPB1*02:01 10
DPB1*03:01 10
DPB1*04:01 52
DPB1*04:02 24

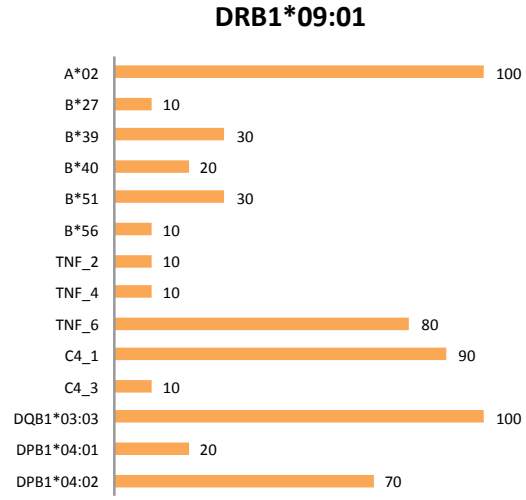

DRB1*04:03
A*02 50
A*32 50
B*27 50
B*40 50
TNF_2 50
TNF_5 25
TNF_7 25
C4_1 75
C4_3 25
BTNL2_3 100
DQB1*03:02 100
DPB1*03:01 25
DPB1*04:01 50
DPB1*05:01 25

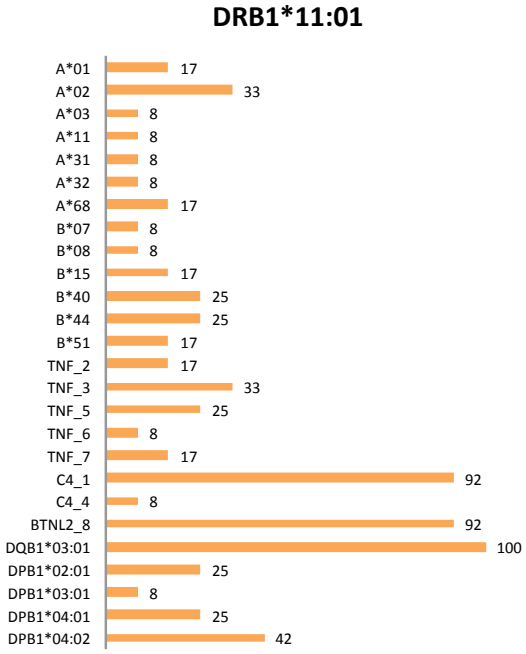

DRB1*04:08
A*02 83
A*03 17
B*27 83
TNF_1 83
TNF_6 17
C4_1 100
BTNL2_3 100
DQB1*03:01 100
DPB1*02:01 83

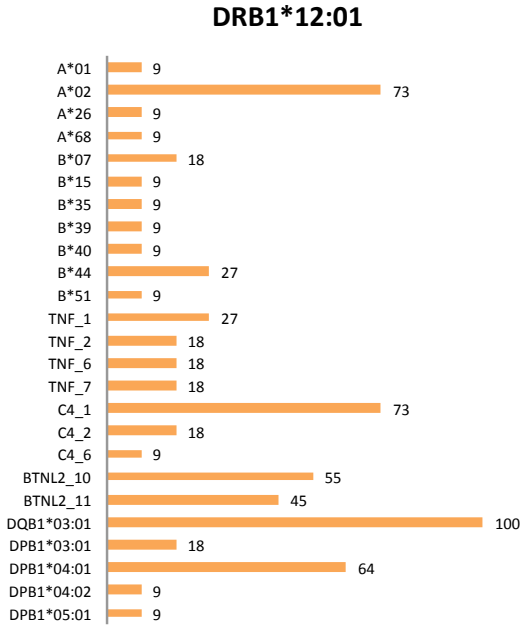

DRB1*04:04
A*02 50
A*03 17
A*29 17
A*32 17
B*15 17
B*27 17
B*44 17
B*51 33
TNF_1 17
TNF_2 17
TNF_7 17
TNF_8 33
C4_1 100
BTNL2_3 50
DQB1*03:02 100
DPB1*02:01 50
DPB1*04:01 50

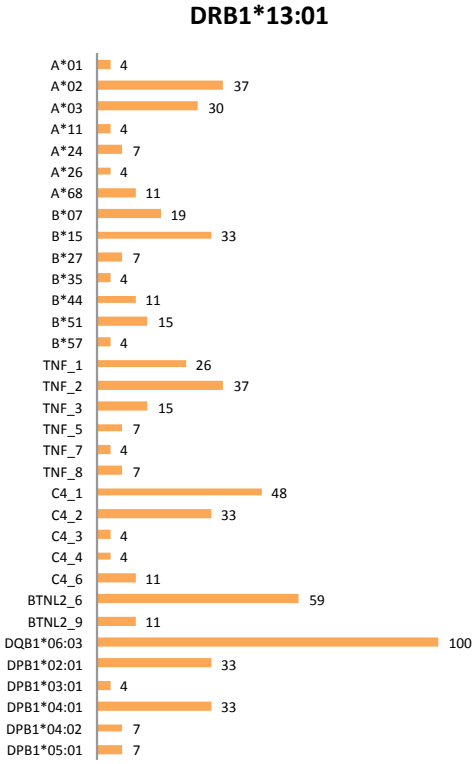

DRB1*07:01
A*01 6
A*02 38
A*03 25
A*24 13
A*26 6
A*29 13
B*07 6
B*13 38
B*15 13
B*35 6
B*40 6
B*44 13
B*57 13
TNF_1 69
TNF_2 6
TNF_3 6
C4_1 88
C4_4 6
BTNL2_3 44
BTNL2_12 25
DQB1*02 75
DQB1*03:01 6
DQB1*03:03 19
DPB1*01:01 6
DPB1*02:01 13
DPB1*04:01 44
DPB1*04:02 25

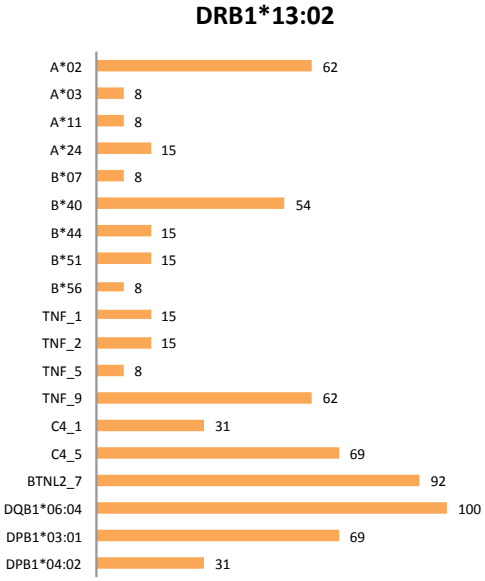

DRB1*08:01
A*01 5
A*02 57
A*03 19
A*24 8
A*68 5
B*08 5
B*15 32
B*27 32
B*39 8
B*40 8
B*44 5
TNF_1 14
TNF_2 41
TNF_3 5
TNF_6 24
TNF_7 5
TNF_8 5
C4_1 68
C4_3 32
BTNL2_4 84
DQB1*04 100
DPB1*03:01 57
DPB1*04:01 30
DPB1*04:02 11


DRB1*09:01
A*02 100
B*27 10
B*39 30
B*40 20
B*51 30
B*56 10
TNF_2 10
TNF_4 10
TNF_6 80
C4_1 90
C4_3 10
DQB1*03:03 100
DPB1*04:01 20
DPB1*04:02 70


DRB1*11:01
A*01 17
A*02 33
A*03 8
A*11 8
A*31 8
A*32 8
A*68 17
B*07 8
B*08 8
B*15 17
B*40 25
B*44 25
B*51 17
TNF_2 17
TNF_3 33
TNF_5 25
TNF_6 8
TNF_7 17
C4_1 92
C4_4 8
BTNL2_8 92
DQB1*03:01 100
DPB1*02:01 25
DPB1*03:01 8
DPB1*04:01 25
DPB1*04:02 42


DRB1*12:01
A*01 9
A*02 73
A*26 9
A*68 9
B*07 18
B*15 9
B*35 9
B*39 9
B*40 9
B*44 27
B*51 9
TNF_1 27
TNF_2 18
TNF_6 18
TNF_7 18
C4_1 73
C4_2 18
C4_6 9
BTNL2_10 55
BTNL2_11 45
DQB1*03:01 100
DPB1*03:01 18
DPB1*04:01 64
DPB1*04:02 9
DPB1*05:01 9


DRB1*13:01
A*01 4
A*02 37
A*03 30
A*11 4
A*24 7
A*26 4
A*68 11
B*07 19
B*15 33
B*27 7
B*35 4
B*44 11
B*51 15
B*57 4
TNF_1 26
TNF_2 37
TNF_3 15
TNF_5 7
TNF_7 4
TNF_8 7
C4_1 48
C4_2 33
C4_3 4
C4_4 4
C4_6 11
BTNL2_6 59
BTNL2_9 11
DQB1*06:03 100
DPB1*02:01 33
DPB1*03:01 4
DPB1*04:01 33
DPB1*04:02 7
DPB1*05:01 7


DRB1*13:02
A*02 62
A*03 8
A*11 8
A*24 15
B*07 8
B*40 54
B*44 15
B*51 15
B*56 8
TNF_1 15
TNF_2 15
TNF_5 8
TNF_9 62
C4_1 31
C4_5 69
BTNL2_7 92
DQB1*06:04 100
DPB1*03:01 69
DPB1*04:02 31

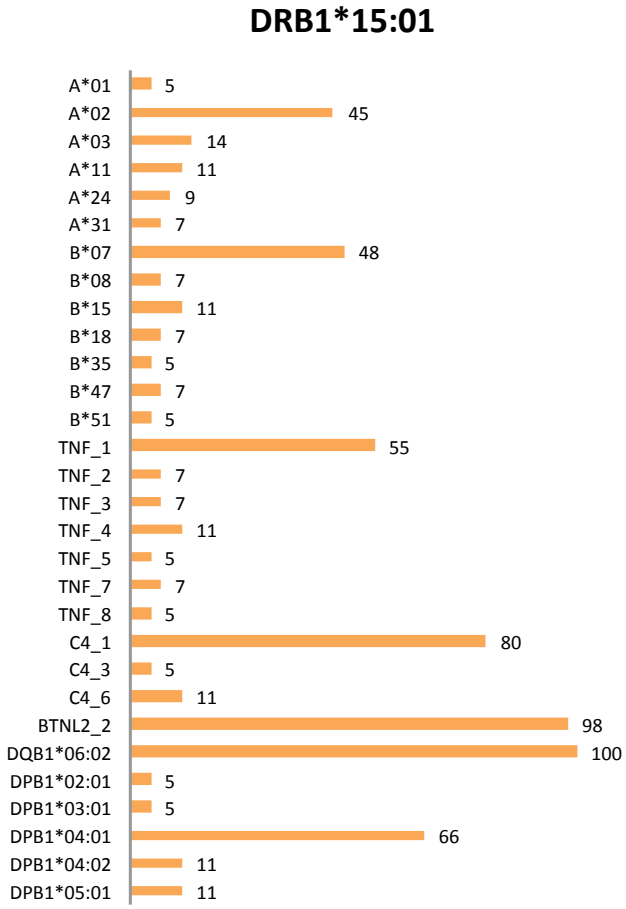


Figure 22 The MHC profile in the Finnish population presented as two-locus haplotypes in relation to *HLA-DRB1*. The frequency (%) of *HLA-DRB1* haplotypes (observed > 3) and traditional HLA alleles (*HLA-A, -B, -DQB1* or *-DPB1*) or MHC class III blocks (TNF, C4 and BTNL2; observed n > 3) shows that the allelic diversity varies in relation to different *HLA-DRB1* alleles. The haplotype map illustrates, for example, that the *HLA-DRB1*15:01* allele is typically linked with *HLA-A*02, HLA-B*07*, TNF_1 block, complement C4_1 block, BTNL2_2 block, *HLA-DQB1*06:02* and/or *HLA-DPB1*04:01*.

6.2 THE MHC ASSOCIATION WITH SARCOIDOSIS

This study confirmed the complex network of effects of several MHC genes to sarcoidosis susceptibility [164]. Based on the results, we suggest a genetically heterogeneous nature for sarcoidosis, showing that both locus heterogeneity (e.g. many variants in BTNL2 associated with sarcoidosis) and allelic heterogeneity (different genes associated with sarcoidosis), which can express the same disease phenotype. We succeeded in replicating the association of variants in HLA-DRB1 and BTNL2 (rs2076530) and detected novel variants in HLA-DPB1 that were associated with sarcoidosis or/and the disease course of sarcoidosis in the Finnish sample. In addition, SNP markers in HLA-DRA and BTNL2, independent of HLA-DRB1, were

significantly associated with sarcoidosis and shared among different populations in a joint case-control study of four European study samples.

Previous studies have highlighted the importance of characterizing the patients into distinct clinical subgroups to avoid inconsistency between studies (e.g., [211,224]). Here, to investigate some of this phenotypic heterogeneity in sarcoidosis, we used a well-characterized patient group that had undergone follow-up for a mean of 15 years and where the patients were divided into subgroups based on the disease activity at a follow-up visit after the first two years. In addition, we collected additional European samples with equally extensive disease phenotyping, to explore variants that were shared among the populations in order to better understand the mechanism of sarcoidosis immunopathogenesis and increase the power of the study. Indeed, environmental triggers are undoubtedly needed for predisposition to sarcoidosis [2], but unfortunately no specific exposure data was available in this study.

Due to the lack of validated population-specific biomarkers in Finland, the prognosis of sarcoidosis is currently based solely on the clinical picture [180]. In Sweden, the *HLA-DRB1*03* allele is used in screening of patients with favourable outcome [192]. If the disease course of sarcoidosis could be estimated with gene susceptibility markers, patients with a risk to develop the chronic form could be recognized at an early stage and treated better. In practice, if a patient has a genetic marker of good prognosis, the treatment may not be initiated in hope of a spontaneous remission. The genetic markers of poor prognosis would promote proper targeting of follow-up visits and early treatment.

6.2.1 MHC CLASS II AND SARCOIDOSIS

In this study, the MHC class II genes were investigated as susceptibility markers for sarcoidosis, because: (i) *HLA-DRB1* has been shown to associate with sarcoidosis in many populations and with distinct disease phenotypes, and had not been studied in the Finnish sample, (ii) the studies of *HLA-DPB1* in sarcoidosis susceptibility have given contradictory results, and (iii) the HLA class II genes encode the molecules on the surface of APCs that bind and present antigens to the T cells. Most importantly, we agreed that for resolved sarcoidosis, variants in the HLA-DRB1 gene, especially HLA-DRB1*03:01 (allelic association and association with peptide-binding sites), are essential for disease predisposition. However, HLA-DRB1*03:01 allele is not as common in Finnish patients as it is among their Swedish counterparts, with carrier frequencies of 19% and 29%, respectively (unpublished results from Study IV). An increasing number of studies has confirmed the importance of HLA-DRB1*03:01 on LS and resolved sarcoidosis susceptibility in patients with European descent [224,297-300]. However, in Japan the HLA-DRB1*03 allele is rare, potentially explaining the non-association of the variant with sarcoidosis [183]. In the Finnish patients, the

HLA-DRB1*15:01 allele did not show association with persistent sarcoidosis, as previously observed in various populations of European descent [211,224,229,301]. Thus, it is perhaps reasonable to hypothesize that the HLA-DRB1*15:01 should not to be used for estimating the prognosis of chronic sarcoidosis in the Finnish population. Furthermore, the protective MHC class II alleles (HLA-DRB1*01:01 and HLA-DPB1*04:02) have shown evidence of association with other autoimmune or infectious diseases as well: HLA-DRB1*01 predisposes for recurrent lymphocytic meningitis, coronary artery disease and juvenile idiopathic arthritis, while HLA-DPB1*04:02 has been shown to be a protective marker for T1D [155,278,302-304]. For chronic sarcoidosis, we did not discover any predisposing HLA-DRB1 or –DPB1 alleles; however, the lack of protective alleles, especially in HLA-DPB1*04:02, shed light on the importance of MHC class II on disease predisposition. Moreover, HLA-DPB1 is strongly associated with CBD, a granulomatous disease clinically and pathophysiologically similar to chronic sarcoidosis [219,305].

In the meta-analysis, some genotyped variants near the *HLA-DRA* promoter region were associated independently of *HLA-DRB1* with sarcoidosis or its disease course. The noncoding region of *HLA-DRA* contains important regulatory elements, including a promoter region and enhancer sequences, which can modify the expression of the protein. The *HLA-DRA* association to sarcoidosis has been shown previously, but the variants detected in Study IV were novel. In addition, *HLA-DRA* and another sarcoidosis susceptibility gene, *ANXAII,* have been suggested to have gene-gene interactions [237,242,244]. To explore the association more thoroughly, studies with dense SNP sets in large samples from multiple ethnic groups are warranted.

6.2.2 BTNL2 AND SARCOIDOSIS

In this Finnish sarcoidosis study, we replicated the association with the exonic missense variant of BTNL2 (rs2076530) [175] with a poor outcome of sarcoidosis, in agreement with previous studies from German, Danish, British, Dutch study samples [175,233,234,306]. Furthermore, the high-risk A-allele of the splice-site variant (rs2076530) results in a truncated protein, which has atypical membrane localization [175,307]. BTNL2 is expressed in APCs, e.g. dendritic cells in the spleen and lymph nodes [308], and is involved in T-cell activation. BTNL2 regulates the immune response by inhibiting T cell proliferation and reducing the production of proinflammatory cytokines. Hence, the dysfunction of BTNL2 could result in an uncontrolled activation of T cells, which is observed in sarcoidosis [180,232,308].

The SNP rs2076530 is probably the most studied variant of the *BTNL2* gene. However, other variants of *BTNL2* gene have been associated with sarcoidosis as well [230,309]. In this study, the finemapping of the *BTNL2*

gene revealed additional variants in the *BTNL2* promoter region, which were significantly associated with different sarcoidosis phenotypes in the meta-analysis combining Finnish, Swedish, Dutch and Czech samples. In addition, population-specific *BTNL2* associations with sarcoidosis phenotypes were found (e.g., rs28362677). In the meta-analysis, the rs2076530 association with sarcoidosis vanished after population stratification and adjusting for *HLA-DRB1* alleles, most probably due to the observed strong LD between *BTNL2* and *HLA-DRB1*. Indeed, the Swedish sample produced a strong LD pattern in MHC class II region, which probably dominated the meta-analysis, leading to a lack of association to rs2076530. Furthermore, several studies have attempted to distinguish whether *BTNL2* or *HLA-DRB1* is responsible for the association to sarcoidosis. A recent Portuguese study [299] showed a phenotype-specific association of *BTNL2* variant rs2076530 in patients without Löfgren syndrome (NL sarcoidosis). Thus they reported that the association was secondary to *HLA-DRB1* in patients with Löfgren syndrome. Again, the associations of *BTNL2* with other autoimmune diseases, including T1D [235] and MS [301], have been shown to be secondary to other genes, mostly the HLA class II alleles. This strongly supports the importance of genotyping *HLA-DRB1*, and adjusting the *BTNL2* associations with HLA class II alleles [192,299]. We may speculate that the discrepancies in the association between *BTNL2* variant rs2076530 and sarcoidosis may be due to heterogeneity in the distributions of alleles in different populations or differences in the study design [175,220,230,233,299,307].

In summary, the importance of BTNL2 in sarcoidosis predisposition is robust. The dysfunction of BTNL2 protein has is a biologically plausible mechanism for sarcoidosis (impairing normal T-cell regulation and response to antigens) and variants of BTNL2 has been associated with different sarcoidosis phenotypes in different populations. In this study, we not only replicated the association of *BTNL2* variant rs2076530 with persistent sarcoidosis in Finns, but our study highlighted the importance of *BTNL2* in sarcoidosis susceptibility in general. Based on the comprehensive analysis of *BTNL2* as sarcoidosis susceptibility variant, we indicated that the combination of modified T cell regulation due to *BTNL2* (e.g., splice-site *BTNL2* variant rs2076530) and the lack of the protective MHC class II alleles (antigen-presentation) are needed for susceptibility to chronic sarcoidosis [226,233,306,310]. For resolving sarcoidosis, we suggest focusing on exploring the widespread variation in the *BTNL2* promoter region and multi-gene haplotypes including *BTNL2*. It is tempting to speculate that different variants of *BTNL2* are present only in certain populations (locus heterogeneity), which suggests that different *BTNL2* genotypes can be expressed as the same disease phenotype, contributing to the lack of replication of the rs2076530 association in some studies.

To conclude, this study illustrated the power of combining well-characterized patients with specialized MHC analysis. However, to validate the role of the MHC variants for disease prognosis of sarcoidosis, the genetic

signals should be confirmed in larger samples and using different techniques, before they can be used in a clinical setting in Finland. For example, it is essential to investigate the expression of the BTNL2 transcript in target cells including the disease-associated missense SNP rs28362677 or study the effect of the promoter SNPs in BTNL2 expression. As there is no well-established animal models for sarcoidosis and as the disease is rather rare disease in the European population, collaboration studies are needed to obtain larger samples sizes, and thus strongly encouraged.

6.3 THE EXTENDED MHC HAPLOTYPE ANALYSIS IS ADVANTAGEOUS FOR COMPLEX DISEASE STUDIES

For many autoimmune and infectious traits, among them sarcoidosis, the strong linkage/association with MHC was discovered decades ago. A typical feature for HLA association is that the any associated variant is common in healthy subjects. For example, the general predisposing allele for sarcoidosis, *HLA-DRB1*15:01,* is the most common allele in the Finnish population (I, II); approximately 30 % of the Finnish sample has at least one copy of *HLA-DRB1*15:01,* and the same allele can act both as a predisposing (for narcolepsy, sarcoidosis and MS) and a protecting (for T1D) allele [311]. Interestingly, multiple autoimmune diseases tend to track in the same patients and same families. Obviously, contribution of additional genes and the poorly understood environmental triggers are necessary for disease predisposition [312].

To understand the genetic and functional basis of the susceptibility variant(s) in MHC, approaches should be employed. As a primary finding, this study showed that by genotyping multiple genes in MHC and establishing the extended MHC haplotypes, additional predisposing/protective associations between markers and trait not detectable with a single MHC allele could be detected (I, II, III, IV). Firstly, *HLA-DRB1*04:01* was not itself associated with the disease course of sarcoidosis, but haplotype *HLA-DRB1*04:01-DPB1*04:01* was (III). Without the meta- and multi-gene analysis, the independent effects of novel *BTNL2* and *HLA-DRA* variants could not have been observed due to strong LD between *BTNL2* and *HLA-DRB1* in Finland (I, IV). Secondly, the *HLA-DRB1*03:01* allele, not the AH 8.1 haplotype [1], was suggested as the most prominent causal variant for resolving sarcoidosis, and would not have been properly identified without genotyping multiple genes (II). Thirdly, different extended MHC haplotypes can lead to functional similarities, including C4 deficiency or a truncated form of *BTNL2,* which can be informative for disease association studies (I) [175,258,283,284].

An additional important aspect in MHC studies is to understand the possible functionality of the associating variant for the trait. Polymorphisms

in the antigen binding sites of HLA genes define the antigen(s) that can be presented in T cells. Variation in the peptide-binding sites of MHC molecules may alter the peptide-MHC interactions and determine which peptides can be bound in the first place [313-315]. It is essential to recognize that similar binding-ability can be observed with different MHC alleles. In our sarcoidosis study, the peptide-binding analysis of HLA molecules clearly indicated that the allelic association of HLA genes does not explain everything, suggesting that it is highly recommended to explore the peptide-binding repertoire of the HLA alleles that can be shared between different alleles, as shown with previous autoimmune studies with, e.g., RA and the shared epitope of *HLA-DRB1* [149,255,316]. Although the primary *HLA-DRB1* associations with sarcoidosis vary between studies and with different populations [187,211,215], the susceptibility/predisposing alleles can affect identical peptide binding sites that vary the ability to bind specific antigens [215,314,317,318]. The peptide binding pockets can have distinct chemical and size characteristics, but any polymorphism of the peptide-binding site can still change the T-cell recognition or have an effect on the peptide binding affinity [255]. In the case of sarcoidosis, the specific antigen is still unknown, and the next challenge is to identify the antigens that the molecules present to CD4+ T cells. It is possible that the antigens differ across distinct sarcoidosis phenotypes or ethnicities [187,211]. The discovery of sarcoidosis-associated antigens could improve the therapy for sarcoidosis.

In addition, many MHC association studies have suggested susceptibility markers for autoimmune and inflammatory traits that are located outside the peptide encoding sequence, and can thus affect epigenetic factors, such as microRNAs that modulate the immune response, methylation sites or binding sites of transcription factors. These studies require more sophisticated analyses [25,168]. The non-replication of variant can be caused by, e.g., population-specific differences, false negatives due to small sample size, heterogeneous replication samples and complex LD structure [233,319]. Our results agree with recent studies showing that the recombination hotspots in MHC vary between populations [2,273,286,287]. Considering the hampering effect of LD, it is highly possible that not all published MHC associations with sarcoidosis are true. The power of these genetic studies can be increased by increasing the size of the study sample, as done in Study IV, combined with accurate phenotyping to lower sample heterogeneity [187].

6.4 CHALLENGES IN MHC ANALYSIS

A comprehensive analysis of the MHC region still remains a daunting challenge due to structural rearrangements, pseudogenes and the LD structure**.** Typical MHC datasets contain a variety of types of genotype information (multiallele HLA genes, SNP or gene copy number data) from different laboratories, with varying nomenclature and genotyping methods

[265,320]. The HLA allele calling and resolution of multiple ambiguous alleles and genotypes requires expertise in both the MHC locus and the population in question, and if the data have not been generated and analyzed properly, ambiguity can lead to major inconsistencies between datasets and spawn false results [321]. In this study, the MHC genotypes were obtained using five different genotyping approaches. To avoid inconsistency in the data, all the genotypes were corroborated by at least two researchers, and the HLA alleles were called with the same nomenclature. In Study IV, we received heterogeneous typing results from three different countries, which were then standardized by converting all the allele data to the lowest resolution (2-digit) [85,293]. In Study IV, we collected additional international samples with mutual extensive disease phenotyping, and performed a successful meta-analysis discovering variants that were shared between the populations, to assist us to better understand the mechanism of sarcoidosis immunopathogenesis.

For discovering novel susceptibility loci associating with the trait, especially outside the MHC region, GWAS offer a cost-effective model by testing a large amount of variants in a large number of patients [82]. In the case of small sample sizes and/or heterogeneous samples, GWASs have had trouble with low statistical power to discover rare variants or with multi-allelic data [212,237]. Our work aimed to study solely the MHC region alone, and replicate the knowm MHC variant(s) in an independent sample (the Finnish study sample) using a functional candidate gene case-control approach. The candidate gene studies are biased, however, as they require prior information regarding gene function and gene variations, and are sensitive to population stratification. Here, the best functional candidate genes in the MHC region were selected using both previous publications and the MHC haplotypic profile of the Finns (Study I).

Finding the proper analysis method for immunogenomic data is tricky, as most of the available statistical approaches (e.g. PLINK [257]) were originally developed to deal with biallelic polymorphism (SNPs). The immunogenetic data warrants tools that can handle the extensive LD within MHC and multiple genes with multiple alleles. To date, there are no single tools available that can perform all the analyses needed for immunogenetic data. First, due to the complexity of the region, the HLA allele calling needs to be performed with specific software, typically a commercial one (e.g., ASSING, SCORE). Second, to conduct MHC haplotypes analysis and calculate the LD between loci, further specialist tools, e.g., PHASE [263] and Arlequin [260], are needed.

The previous studies have highlighted several issues that should be addressed in immunogenetic data analysis, including the cut-off values for low-frequency variants and the caution needed in interpreting LD between loci [84,104,112,265,321]. As suggested elsewhere [265], analytical interpretations should not rely on rare haplotypes. In this study, we used a haplotype frequency cut-off value, excluding haplotypes with frequency less

than 1%. Although possible false results were thus rejected, many potentially informative low-frequency haplotypes that may be important for disease association studies were excluded as well [265,293]. LD complicates inding the causal variant as many markers are in total LD ($r^2=1$) while genes in conversed haplotypes (e.g., *HLA-DRB1*15:01* is always with *HLA-DQB1*06:02* in the Finnish population) may have equivalent statistical proof of association. Moreover, the linkage and association signals can cover long haplotypes with alleles from multiple HLA genes [2,82]. As several different identifiers may exist for a single variant, this complicates comparing variants across studies [321]. The improper interpretation of LD (e.g., D', is sensitive for rare alleles and r^2 is not suitable for multi-allelic data analysis [265]) can create misleading results, and thus it is important to explore results using a variety of LD measures and take into account the allele counts, as we did in Study I. Again, the strong LD within MHC and the heterogeneity of sarcoidosis phenotypes complicates the interpretation and detection of causal variants. Most importantly, in many studies the degree of LD with HLA class II alleles and adjusting for disease phenotypes have not been stratified [306], potentially leading to discrepancies between studies [299]. Furthermore, the adjustments can change the association with the reported variants, as shown in the study of Spangolo et al [233], where the rs2076530 association with sarcoidosis vanished after adjusting for *HLA-DRB1* alleles and exclusion of genetically district disease group (LS). To note, Valentonyte et al. [175] did not exclude the patients with Löfgren syndrome from their studies.

The current HLA typing methods are targeted approaches, requiring the amplification of specific HLA gene segments. With innovative NGC techniques (whole genome, whole exome and transcriptome sequencing), multiple HLA genes can be sequenced at once while preserving phase information. However, the complexity of MHC warrants bioinformatics, scarcity of population-specific references and knowledge of the MHC region, and most likely commercial allele calling software, that are not widely available. [84,85,95,320]. One of the first publications using NGS data for HLA typing was performed with a Roche-454, using long reads from targeted sequencing data [84,95]. A recent study showed higher than 90% accuracy using NGS exome sequence (Illumina) of HLA class I genes [100]. In addition, recent NGS studies have been hampered with insufficient read depth in whole-genome and exome sequencing, highlighting similar MHC-specific challenges in NGS than in traditional Sanger sequencing; (i) most subjects are heterozygous in MHC locus, making the alignment with the reference challenging, (ii) to date, thousands of known HLA alleles have been identified and, (iii) MHC region contains long, segmental duplications, heterodimeric proteins with several possible genes and pseudogenes [84,85,95,97,100-102,322]. Previously, a tag-SNP based HLA-typing method was suggested as an alternative solution for traditional HLA typing [82,86]. In some cases, the HLA tag-SNPs have been validated in a specific

population and used successfully instead of traditional HLA genotyping [89]. Due to the complexity of MHC region, the allele differences between populations and the lack of population-specific reference data complicates the tag-SNPing and can lead to false imputation results [27,82,323]. The SNPs in this study were initially selected to cover the genes in question, not to impute HLA alleles. However, we had the *HLA-DRB1*15:01* tag-SNP present (rs3135388, r^2=0.966, D'=0.993[86]), and found that in the Finnish sample, the LD between *HLA-DRB1*15:01* and rs3135388 was complete (r^2 =1.00 and D'= 1.0). The SNP has been used to impute *HLA-DRB1*15:01* in previous studies [282], but as far as we know no validation in a Finnish sample has been performed. Despite of the success of *HLA-DRB1*15* imputation, based on our unpublished data, the current SNP imputation of MHC markers is not suitable for detecting all polymorphism in the MHC region, especially in Finland. To detect Finnish tag-SNPs for HLA alleles, a large population sample with several HLA genotypes and dense SNP fine-mapping in the MHC region is required, and this was not available at the time of our study.

7 CONCLUDING REMARKS AND FUTURE PROSPECTS

The first MHC was defined in mice over 70 years ago [324], and the first HLA antigen MAC (later to become HLA-A2) was identified over 50 years ago [325]. The first correlation between HLA matching and kidney allograft survival was shown in 1965 and the first remarkable HLA disease association was detected between HLA-B27 and ankylosing spondylitis in 1972 [141,326]. Since then, the study of MHC genes in different species and in ethnic populations has dramatically increased, leading to many more identified MHC disease associations [84,327].

Despite the genetic and epidemiological efforts, understanding the genetic basis of many MHC associated traits, such as sarcoidosis, still remains a mystery [82,176]. Furthermore, while the MHC genes have been the most extensively studied susceptibility genes in sarcoidosis, Finnish MHC studies of sarcoidosis did not exist prior to this study. Although the traditional HLA haplotype frequencies for the Finns have been published previously [288,291], the diversity of the extended MHC haplotypes had not been studied among the Finns. We hypothesized that the study of the Finns MHC profile and the extended MHC structure would be advantageous for pinpointing possible causal variant(s) for sarcoidosis.

After analysing several HLA and non-HLA genes in Finnish subjects and SNPs in sarcoidosis collaboration study subjects, this study succeeded in the following five goals.

1) The characterization of the MHC profile in Finnish population, using extended MHC haplotypes.
2) The replication of known sarcoidosis-associated MHC markers in a Finnish sample
3) The detection of novel associations between sarcoidosis phenotypes and MHC markers that were shared among populations
4) The identification of novel population-specific MHC associations for sarcoidosis phenotypes
5) Demonstration of guidelines for handling immunogenetic data

The Finnish MHC profiles revealed certain population-specific MHC haplotypes with a length and LD structure that a unique in Finns. The study of extended MHC haplotypes enabled the creation of a two-locus linkage map in relation to *HLA-DRB1* alleles, which illustrates a linkage between the two variants. In addition, distinct extended MHC haplotypes can lead to functional similarities, including C4 deficiency or a truncated form of *BTNL2* [175]. Thus, the rare *HLA-DRB1* haplotypes that differ from the common haplotypes by, e.g., a particular *TNF* polymorphism, are potentially informative for future association studies.

All the main achievements in the sarcoidosis study promoted the importance of the MHC in sarcoidosis predisposition. Even though the Finnish sample in this study was limited in size, our study provides new insights to the disease prognosis, highlighting a particular distribution of MHC markers with some markers associated with different patterns of disease progression. For sarcoidosis, an immune-mediated inflammatory disease, the proper understanding of the complex etiology warrants further studies of the MHC in different ethnic groups, preferably via a meta-analysis of samples with well-characterized disease phenotypes. In addition, the non-MHC variants that have shown to associate with sarcoidosis phenotypes (detected through GWASs) ought to be investigated in a Finnish study sample. Biologically, the most interesting of these candidate genes is *annexin A11* on 10q22.3[242].

Similarily to previously published work [84], this study agreed that the major challenges in MHC research are the complexity of the MHC region, including strong LD, heterogeneity of associated disease phenotypes, and population-specific MHC allele distribution. This strongly supports the importance of genotyping multiple genes in the MHC region and using the extended haplotype analysis for detecting causal variants for MHC-associated diseases. Using this approach, novel sarcoidosis-associated MHC variants were identified, which would have otherwise been undetectable.

To summarize, the MHC region offers considerable potential to discover predisposing and protective variants affecting autoimmune and infectious disease [168]. Analysis of the MHC region, however, is often ignored in genetic studies. MHC genetics still requires special expertise to account for the special challenges involved. The new MHC genotyping approaches, e.g., next-generation sequencing of exomes and transcriptomes and population-specific SNP-tagging of MHC alleles, will likely offer a more straightforward and cost-effective methods for MHC analyses, especially for those not familiar with the fascinating MHC region [82,84,95]. It is highly plausible that this leads to considerable advancements in understanding the region and its associated traits in the years to come.

8 ACKNOWLEDGEMENTS

This study was carried out in the MHC Research Group at the Transplantation Laboratory, Haartman Institute, University of Helsinki. I wish to express my deepest gratitude to the former and current head of the Transplantation Laboratory Professor Pekka Häyry and Professor Risto Renkonen, and the former and current Head of the Haartman Institute, Professor Seppo Meri and Professor Tom Böhling, for providing excellent research facilities and an encouraging scientific research environment.

I wish to thank all those study participants, who have made it possible to conduct this study. This Thesis work was a team effort.

I want to sincerely acknowledge all the funding sources for making this Thesis possible. Financial support was provided by Nummela foundation, Helsinki Biomedical Graduate School (HBGP and LERU), HLA-NET, EFI, University of Helsinki Funds and Hengitysliitto.

I wish to express my heartfelt appreciation to my supervisor Docent Marja-Liisa "Maisa" Lokki. I want to thank you for letting me to have my space and freedom to do research. You have always believed in me, even when I was totally lost. I admire your knowledge of MHC, and I thank you for sharing that "secret" with me. I owe a large debt of gratitude to the wonderful clinicians, Professor Olof Selroos and Docent Anne Pietinalho. This work could not have been possible without your guidance and clinical support that you have given me. In addition, I wish to thank all the other clinicians who have made this study possible together with the patients.

I sincerely wish to thank Docent Janna Saarela for accepting the invitation to be the opponent at my thesis defense. Professor Karl "Kalle" Lemström is warmly thanked for accepting the role of custos (and taking care of my belly-button i.e., surgical wounds after appendectomy). Professors Hannes Lohi and Pentti Tienari are thanked for accepting the role of pre-examiners and their constructive suggestions to improve the manuscript. I thank you also for working under a tight schedule. Hannes is also thanked for carrying the Halloween pumpkins. Leena Saraste and Verneri Anttila are thanked for their wonderful scientific support and excellent language check of the thesis. Leena is also thanked for helping me in numerous practical matters.

The members of my Thesis committee Mari Kaunisto and Professor Tarja Laitinen are gratefully acknowledged for their support and contribution during this work. I enjoyed our conversations and providing valuable comments and fruitful criticism.

I wish to thank all of my co-authors, not mentioned before, to whom I am grateful and indebted: Johan Grunewald, C.H.M van Moorsel, Martin Petrek Anders Eklund, J.C Grutters, V.Kolek, F.Mrazek, L.Padyukov, M.Ronninger,

Mikko Seppänen, Anil Palikhe, Hanna Vauhkonen, Katja Eronen, Ida Surakka, Verneri Anttila, Effie Vlachopoulou, Elisa Lahtela, Riitta Paakkanen, Jouni Hedman, Minna Purokivi, E. Varkki, Jagoda Lasota and Krista Salli for the time, effort and work that you have put into my studies. This thesis would not have been possible without the great collaboration that we had.

I thank from the bottom of my heart my first mentors in genetics: Docent Maija Wessman, Docent Arto Orpana and Professors Leena and Aarno Palotie, not forgetting Eija Hämäläinen. At the age of 16 in 1998 you took me under your guidance and showed me what science really is. I also want to thank all the former co-workers in the Finnish Genome Center (2003-2007; Docent Elisabeth Widen, Docent Päivi Lahermo and others, especially Anu, Sirkku, Susanna, Annika, Jouko, Timo and Virpi) and the Migraine Group (2004-2007; Aarno, Maija, Mikko, Ville, Päivi, Mari and Verneri, and MSc thesis supervisor Teppo Varilo) for their support and expertise in genetics. The migraine girls (Mari and Päivi) have a special place in my heart. Verneri is also thanked for sharing his apartment in Boston and letting me to finish my Thesis (on his couch). Go Socks!

This Thesis is dedicated to my wonderful colleagues, old and new. A special thanks go to our Motivation group: Ville, Hilkka, Sakari and Nina. Ville, we made it! The endless conversations over sweets and sushi made the whole life better. Eeva, I will never forget our Pulla Battle. Even though you won, I still think my pulla is the best. Elisa, you have always been there for me- helping me with the wet lab or bringing me some sweets. We also had similar desks with piles of paper. Antti (the new Ville) is thanked for sharing his ideas and practical things during the last weeks.

I also wish to thank the co-workers from University of Helsinki and THL: Nipsu, Riitta, Efi, Johanna, Eeva, Leena, Heikki, Lauri, Misu, Minna V, Minna B, Kaisa, Lauri, Krista, Jagoda, Katja, Pia, Anil, Rainer, Maria, Marja, Sami, Hannat, Kalle's boys, Virpi, Johannes (Furbyy), Tiia, Emmi, PP, Ida, Mitja, Olli, Jarkko, Mikko, Annu, Hanna, Justin Bieber, Miisa, SampoBisnari, Tero…and those not mentioned here. I thank you all for constructive and valuable comments during these years. Also, we had the best Christmas parties and conference trips.

Very special thanks go to Professor Markus Perola and his superb group (Tero, Anni, Perttu, Marjis, Kirsi, Natalie, Iiro, Hannele and Outi). Our journey has just begun!

I have been exceptionally lucky to have a number of loving friends in my life: Laura "Kastis", Olli, Laura "Helis", Annin(k)a, Sampo, Iiris, Pala, Anne, Jon, Johanna, Tuomas, Annasukava, Vassiset, Jonna, Radiolinja girls (Suvi and Krisse), Katarina, Laajasalo girls (Janica and Anni), KT&Essi, Leo&Nora, Jaakko, Samuli&Emppu, Fille, The Team Florida 2013 (Olli&Wilma), The Sushibar team (Matti&Andu)…. I also want to thank all the other boys from Sisäpiiri and Rähinä Dösäjengi for keeping my husband busy, especially during these last few months. Thank you all so much!

I have written a huge part of this Thesis in the Taivallahti Tennis club. A very special thanks go to my Taivis Tennis team: coaches Anders, Janne, Jermo and Reijo, masseur Levo and playmates Ruusu, Erkka, Heidi, Suski, Ellu, Nicce, and many others. Thank you all for your cheerful company and friendship.

My warmest thanks go to my family. I want to thank my parents, Marketta and Pekka, for their unconditional love and support throughout my life. I thank my brother Patrik and his wife Katja (and her family), my dear grandma Mami, my family-in-law (Päivi ”Pivi”, Johan “Jukepike”, Ape, Jenni and Jussshba), and the rest of the big Family (Hanski, Pessu, Miksu, Joanna, Sampo, Ripa, Wallu, Hannu, Joan, Kristina, Chris, Saara, Jussi, Anja, Sointu, and many others). I am whole heartily thankful to my children, Matilda and Lucas, for letting me to work almost 24-7 for many weeks. I love you so much. You asked me many, many times “Onks se nyt valmis jo?”. NYT ON!

Finally Markus, thank you for always being there for me. Your support has been priceless. You are my best friend and the love of my life.

Annika
15th of January 4:20 AM

9 REFERENCES

1. Price P, Witt C, Allcock R, Sayer D, Garlepp M, et al. (1999) The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. Immunol Rev 167: 257-274.
2. Horton R, Gibson R, Coggill P, Miretti M, Allcock RJ, et al. (2008) Variation analysis and gene annotation of eight MHC haplotypes: The MHC haplotype project. Immunogenetics 60: 1-18. 10.1007/s00251-007-0262-2; 10.1007/s00251-007-0262-2.
3. McDevitt H. (2002) The discovery of linkage between the MHC and genetic control of the immune response. Immunol Rev 185: 78-85.
4. Iannuzzi MC, Rybicki BA. (2007) Genetics of sarcoidosis: Candidate genes and genome scans. Proc Am Thorac Soc 4: 108-116. 10.1513/pats.200607-141JG.
5. Pietinalho A, Ohmichi M, Lofroos AB, Hiraga Y, Selroos O. (2000) The prognosis of pulmonary sarcoidosis in finland and hokkaido, japan. A comparative five-year study of biopsy-proven cases. Sarcoidosis Vasc Diffuse Lung Dis 17: 158-166.
6. Selroos O. (1969) The frequency, clinical picture and prognosis of pulmonary sarcoidosis in finland. Acta Med Scand Suppl 503: 3-73.
7. Muller-Quernheim J, Schurmann M, Hofmann S, Gaede KI, Fischer A, et al. (2008) Genetics of sarcoidosis. Clin Chest Med 29: 391-414, viii. 10.1016/j.ccm.2008.03.007; 10.1016/j.ccm.2008.03.007.
8. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921. 10.1038/35057062.
9. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science 291: 1304-1351. 10.1126/science.1058040.
10. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, et al. (2012) GENCODE: The reference human genome annotation for the ENCODE project. Genome Res 22: 1760-1774. 10.1101/gr.135350.111; 10.1101/gr.135350.111.
11. Hannan AJ. (2012) Tandem repeat polymorphisms: Mediators of genetic plasticity, modulators of biological diversity and dynamic sources of disease susceptibility. Adv Exp Med Biol 769: 1-9.
12. Smale ST, Kadonaga JT. (2003) The RNA polymerase II core promoter. Annu Rev Biochem 72: 449-479. 10.1146/annurev.biochem.72.121801.161520.
13. Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. (2008) The RNA polymerase II core promoter - the gateway to transcription. Curr Opin Cell Biol 20: 253-259. 10.1016/j.ceb.2008.03.003; 10.1016/j.ceb.2008.03.003.
14. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, et al. (2006) Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 38: 626-635. 10.1038/ng1789.
15. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, et al. (2012) Landscape of transcription in human cells. Nature 489: 101-108. 10.1038/nature11233; 10.1038/nature11233.
16. Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, et al. (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature 408: 86-89. 10.1038/35040556.
17. Anastasi G, Cutroneo G, Santoro G, Arco A, Rizzo G, et al. (2008) Costameric proteins in human skeletal muscle during muscular inactivity. J Anat 213: 284-295. 10.1111/j.1469-7580.2008.00921.x; 10.1111/j.1469-7580.2008.00921.x.
18. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, et al. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in caenorhabditis elegans. Nature 403: 901-906. 10.1038/35002607.
19. Gervin K, Vigeland MD, Mattingsdal M, Hammero M, Nygard H, et al. (2012) DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis:

Identification of epigenetically dysregulated genes. PLoS Genet 8: e1002454. 10.1371/journal.pgen.1002454; 10.1371/journal.pgen.1002454.
20. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, et al. (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 325: 1246-1250. 10.1126/science.1174148; 10.1126/science.1174148.
21. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56-65. 10.1038/nature11632; 10.1038/nature11632.
22. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, et al. (2012) The accessible chromatin landscape of the human genome. Nature 489: 75-82. 10.1038/nature11232; 10.1038/nature11232.
23. Howald C, Tanzer A, Chrast J, Kokocinski F, Derrien T, et al. (2012) Combining RT-PCR-seq and RNA-seq to catalog all genic elements encoded in the human genome. Genome Res 22: 1698-1710. 10.1101/gr.134478.111; 10.1101/gr.134478.111.
24. Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, et al. (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. Nature 489: 83-90. 10.1038/nature11212; 10.1038/nature11212.
25. ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, et al. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57-74. 10.1038/nature11247; 10.1038/nature11247.
26. Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, et al. (2012) The 1000 genomes project: Data management and community access. Nat Methods 9: 459-462. 10.1038/nmeth.1974; 10.1038/nmeth.1974.
27. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, et al. (2010) Integrating common and rare genetic variation in diverse human populations. Nature 467: 52-58. 10.1038/nature09298; 10.1038/nature09298.
28. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, et al. (2012) Rate of de novo mutations and the importance of father's age to disease risk. Nature 488: 471-475. 10.1038/nature11396; 10.1038/nature11396.
29. Lynch M. (2010) Evolution of the mutation rate. Trends Genet 26: 345-352. 10.1016/j.tig.2010.05.003; 10.1016/j.tig.2010.05.003.
30. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. (2009) An operational definition of epigenetics. Genes Dev 23: 781-783. 10.1101/gad.1787609; 10.1101/gad.1787609.
31. Koppelman GH, Nawijn MC. (2011) Recent advances in the epigenetics and genomics of asthma. Curr Opin Allergy Clin Immunol 11: 414-419. 10.1097/ACI.0b013e32834a9573; 10.1097/ACI.0b013e32834a9573.
32. Yang IV, Schwartz DA. (2012) Epigenetic mechanisms and the development of asthma. J Allergy Clin Immunol 130: 1243-1255. 10.1016/j.jaci.2012.07.052; 10.1016/j.jaci.2012.07.052.
33. Kempthorne O, Osborne RH. (1961) The interpretation of twin data. Am J Hum Genet 13: 320-339.
34. Hawkes CH. (1997) Twin studies in medicine--what do they tell us? QJM 90: 311-321.
35. Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, et al. (2003) Heritability of adult body height: A comparative study of twin cohorts in eight countries. Twin Res 6: 399-408. 10.1375/136905203770326402.
36. Delves PJ, Roitt IM. (2000) The immune system. first of two parts. N Engl J Med 343: 37-49. 10.1056/NEJM200007063430107.
37. Delves PJ, Roitt IM. (2000) The immune system. second of two parts. N Engl J Med 343: 108-117. 10.1056/NEJM200007133430207.
38. Rock FL, Hardiman G, Timans JC, Kastelein RA, Bazan JF. (1998) A family of human receptors structurally related to drosophila toll. Proc Natl Acad Sci U S A 95: 588-593.
39. Ting JP, Trowsdale J. (2002) Genetic control of MHC class II expression. Cell 109 Suppl: S21-33.
40. Harrington LE, Hatton RD, Mangan PR, Turner H, Murphy TL, et al. (2005) Interleukin 17-producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. Nat Immunol 6: 1123-1132. 10.1038/ni1254.

41. Xiong W, Lahita RG. (2013) Pragmatic approaches to therapy for systemic lupus erythematosus. Nat Rev Rheumatol . 10.1038/nrrheum.2013.157; 10.1038/nrrheum.2013.157.
42. Moulton VR, Tsokos GC. (2011) Abnormalities of T cell signaling in systemic lupus erythematosus. Arthritis Res Ther 13: 207. 10.1186/ar3251; 10.1186/ar3251.
43. Waxman SG. (1998) Demyelinating diseases--new pathological insights, new therapeutic targets. N Engl J Med 338: 323-325. 10.1056/NEJM199801293380610.
44. Jersild C, Dupont B, Fog T, Platz PJ, Svejgaard A. (1975) Histocompatibility determinants in multiple sclerosis. Transplant Rev 22: 148-163.
45. Lehuen A, Diana J, Zaccone P, Cooke A. (2010) Immune cell crosstalk in type 1 diabetes. Nat Rev Immunol 10: 501-513. 10.1038/nri2787; 10.1038/nri2787.
46. Todd JA, Bell JI, McDevitt HO. (1987) HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. Nature 329: 599-604. 10.1038/329599a0.
47. Eisenbarth GS. (2004) Type 1 diabetes: Molecular, cellular and clinical immunology. Adv Exp Med Biol 552: 306-310.
48. Reiner SL, Sallusto F, Lanzavecchia A. (2007) Division of labor with a workforce of one: Challenges in specifying effector and memory T cell fate. Science 317: 622-625. 10.1126/science.1143775.
49. The MHC sequencing consortium. (1999) Complete sequence and gene map of a human major histocompatibility complex. the MHC sequencing consortium. Nature 401: 921-923. 10.1038/44853.
50. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, et al. (2004) Gene map of the extended human MHC. Nat Rev Genet 5: 889-899. 10.1038/nrg1489.
51. Piertney SB, Oliver MK. (2006) The evolutionary ecology of the major histocompatibility complex. Heredity (Edinb) 96: 7-21. 10.1038/sj.hdy.6800724.
52. Prugnolle F, Manica A, Charpentier M, Guegan JF, Guernier V, et al. (2005) Pathogen-driven selection and worldwide HLA class I diversity. Curr Biol 15: 1022-1027. 10.1016/j.cub.2005.04.050.
53. Kasahara M, Nakaya J, Satta Y, Takahata N. (1997) Chromosomal duplication and the emergence of the adaptive immune system. Trends Genet 13: 90-92.
54. Germain RN. (1994) MHC-dependent antigen processing and peptide presentation: Providing ligands for T lymphocyte activation. Cell 76: 287-299.
55. Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, et al. (2003) The DNA sequence and analysis of human chromosome 6. Nature 425: 805-811. 10.1038/nature02055.
56. Thorsby E, Lie BA. (2005) HLA associated genetic predisposition to autoimmune diseases: Genes involved and possible mechanisms. Transpl Immunol 14: 175-182. 10.1016/j.trim.2005.03.021.
57. Smith WP, Vu Q, Li SS, Hansen JA, Zhao LP, et al. (2006) Toward understanding MHC disease associations: Partial resequencing of 46 distinct HLA haplotypes. Genomics 87: 561-571. 10.1016/j.ygeno.2005.11.020.
58. Gussow D, Rein R, Ginjaar I, Hochstenbach F, Seemann G, et al. (1987) The human beta 2-microglobulin gene. primary structure and definition of the transcriptional unit. J Immunol 139: 3132-3138.
59. Gough SC, Simmonds MJ. (2007) The HLA region and autoimmune disease: Associations and mechanisms of action. Curr Genomics 8: 453-465. 10.2174/138920207783591690; 10.2174/138920207783591690.
60. van Bleek GM, Nathenson SG. (1991) The structure of the antigen-binding groove of major histocompatibility complex class I molecules determines specific selection of self-peptides. Proc Natl Acad Sci U S A 88: 11032-11036.
61. Schumacher TN, De Bruijn ML, Vernie LN, Kast WM, Melief CJ, et al. (1991) Peptide selection by MHC class I molecules. Nature 350: 703-706. 10.1038/350703a0.
62. Natarajan K, Li H, Mariuzza RA, Margulies DH. (1999) MHC class I molecules, structure and function. Rev Immunogenet 1: 32-46.
63. Brusic V, Bajic VB, Petrovsky N. (2004) Computational methods for prediction of T-cell epitopes--a framework for modelling, testing, and applications. Methods 34: 436-443. 10.1016/j.ymeth.2004.06.006.

64. Nelson CA, Fremont DH. (1999) Structural principles of MHC class II antigen presentation. Rev Immunogenet 1: 47-59.
65. Dani A, Chaudhry A, Mukherjee P, Rajagopal D, Bhatia S, et al. (2004) The pathway for MHCII-mediated presentation of endogenous proteins involves peptide transport to the endo-lysosomal compartment. J Cell Sci 117: 4219-4230. 10.1242/jcs.01288.
66. Kelley J, Walter L, Trowsdale J. (2005) Comparative genomics of natural killer cell receptor gene clusters. PLoS Genet 1: 129-139. 10.1371/journal.pgen.0010027.
67. Younger RM, Amadou C, Bethel G, Ehlers A, Lindahl KF, et al. (2001) Characterization of clustered MHC-linked olfactory receptor genes in human and mouse. Genome Res 11: 519-530. 10.1101/gr.160301.
68. Gunther E, Walter L. (2001) The major histocompatibility complex of the rat (rattus norvegicus). Immunogenetics 53: 520-542. 10.1007/s002510100361.
69. Wagner JL. (2003) Molecular organization of the canine major histocompatibility complex. J Hered 94: 23-26.
70. Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, et al. (1999) The chicken B locus is a minimal essential major histocompatibility complex. Nature 401: 923-925. 10.1038/44856.
71. Madden DR. (1995) The three-dimensional structure of peptide-MHC complexes. Annu Rev Immunol 13: 587-622. 10.1146/annurev.iy.13.040195.003103.
72. Guillemot F, Billault A, Pourquie O, Behar G, Chausse AM, et al. (1988) A molecular map of the chicken major histocompatibility complex: The class II beta genes are closely linked to the class I genes and the nucleolar organizer. EMBO J 7: 2775-2785.
73. Kasahara M, Watanabe Y, Sumasu M, Nagata T. (2002) A family of MHC class I-like genes located in the vicinity of the mouse leukocyte receptor complex. Proc Natl Acad Sci U S A 99: 13687-13692. 10.1073/pnas.212375299.
74. Watanabe Y, Maruoka T, Walter L, Kasahara M. (2004) Comparative genomics of the mill family: A rapidly evolving MHC class I gene family. Eur J Immunol 34: 1597-1607. 10.1002/eji.200424919.
75. Holling TM, Schooten E, van Den Elsen PJ. (2004) Function and regulation of MHC class II molecules in T-lymphocytes: Of mice and men. Hum Immunol 65: 282-290. 10.1016/j.humimm.2004.01.005.
76. Kuroda N, Figueroa F, O'hUigin C, Klein J. (2002) Evidence that the separation of mhc class II from class I loci in the zebrafish, danio rerio, occurred by translocation. Immunogenetics 54: 418-430. 10.1007/s00251-002-0473-5.
77. Bingulac-Popovic J, Figueroa F, Sato A, Talbot WS, Johnson SL, et al. (1997) Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, danio rerio. Immunogenetics 46: 129-134.
78. Robinson J, Halliwell JA, McWilliam H, Lopez R, Parham P, et al. (2013) The IMGT/HLA database. Nucleic Acids Res 41: D1222-7. 10.1093/nar/gks949; 10.1093/nar/gks949.
79. Kennedy LJ, Barnes A, Short A, Brown JJ, Lester S, et al. (2007) Canine DLA diversity: 1. new alleles and haplotypes. Tissue Antigens 69 Suppl 1: 272-288. 10.1111/j.1399-0039.2006.00779.x.
80. Laforet M, Froelich N, Parissiadis A, Pfeiffer B, Schell A, et al. (1997) A nucleotide insertion in exon 4 is responsible for the absence of expression of an HLA-A*01 allele. Tissue Antigens 50: 347-350.
81. Cano P, Klitz W, Mack SJ, Maiers M, Marsh SG, et al. (2007) Common and well-documented HLA alleles: Report of the ad-hoc committee of the american society for histocompatiblity and immunogenetics. Hum Immunol 68: 392-417. 10.1016/j.humimm.2007.01.014.
82. de Bakker PI, Raychaudhuri S. (2012) Interrogating the major histocompatibility complex with high-throughput genomics. Hum Mol Genet 21: R29-36. 10.1093/hmg/dds384.
83. Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, et al. (2003) IMGT/HLA and IMGT/MHC: Sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res 31: 311-314.
84. Erlich H. (2012) HLA DNA typing: Past, present, and future. Tissue Antigens 80: 1-11. 10.1111/j.1399-0039.2012.01881.x; 10.1111/j.1399-0039.2012.01881.x.
85. Dunn PP. (2011) Human leucocyte antigen typing: Techniques and technology, a critical appraisal. Int J Immunogenet 38: 463-473. 10.1111/j.1744-313X.2011.01040.x; 10.1111/j.1744-313X.2011.01040.x.

86. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, et al. (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet 38: 1166-1172. 10.1038/ng1885.
87. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, et al. (2007) A whole-genome association study of major determinants for host control of HIV-1. Science 317: 944-947. 10.1126/science.1143767.
88. Colombo S, Rauch A, Rotger M, Fellay J, Martinez R, et al. (2008) The HCP5 single-nucleotide polymorphism: A simple screening tool for prediction of hypersensitivity reaction to abacavir. J Infect Dis 198: 864-867. 10.1086/591184; 10.1086/591184.
89. Monsuur AJ, de Bakker PI, Zhernakova A, Pinto D, Verduijn W, et al. (2008) Effective detection of human leukocyte antigen risk alleles in celiac disease using tag single nucleotide polymorphisms. PLoS One 3: e2270. 10.1371/journal.pone.0002270; 10.1371/journal.pone.0002270.
90. Dilthey AT, Moutsianas L, Leslie S, McVean G. (2011) HLA*IMP--an integrated framework for imputing classical HLA alleles from SNP genotypes. Bioinformatics 27: 968-972. 10.1093/bioinformatics/btr061; 10.1093/bioinformatics/btr061.
91. Leslie S, Donnelly P, McVean G. (2008) A statistical method for predicting classical HLA alleles from SNP data. Am J Hum Genet 82: 48-56. 10.1016/j.ajhg.2007.09.001; 10.1016/j.ajhg.2007.09.001.
92. Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, et al. (2013) HIBAG-HLA genotype imputation with attribute bagging. Pharmacogenomics J . 10.1038/tpj.2013.18; 10.1038/tpj.2013.18.
93. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, et al. (2013) Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One 8: e64683. 10.1371/journal.pone.0064683; 10.1371/journal.pone.0064683.
94. Koskinen L, Romanos J, Kaukinen K, Mustalahti K, Korponay-Szabo I, et al. (2009) Cost-effective HLA typing with tagging SNPs predicts celiac disease risk haplotypes in the finnish, hungarian, and italian populations. Immunogenetics 61: 247-256. 10.1007/s00251-009-0361-3; 10.1007/s00251-009-0361-3.
95. De Santis D, Dinauer D, Duke J, Erlich HA, Holcomb CL, et al. (2013) 16(th) IHIW : Review of HLA typing by NGS. Int J Immunogenet 40: 72-76. 10.1111/iji.12024; 10.1111/iji.12024.
96. Lee H, Tang H. (2012) Next-generation sequencing technologies and fragment assembly algorithms. Methods Mol Biol 855: 155-174. 10.1007/978-1-61779-582-4_5; 10.1007/978-1-61779-582-4_5.
97. Iqbal Z, Neveling K, Razzaq A, Shahzad M, Zahoor MY, et al. (2012) Targeted next generation sequencing reveals a novel intragenic deletion of the TPO gene in a family with intellectual disability. Arch Med Res 43: 312-316. 10.1016/j.arcmed.2012.01.011; 10.1016/j.arcmed.2012.01.011.
98. Feuk L, Carson AR, Scherer SW. (2006) Structural variation in the human genome. Nat Rev Genet 7: 85-97. 10.1038/nrg1767.
99. Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, et al. (2012) Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol 30: 771-776. 10.1038/nbt.2303.
100. Major E, Rigo K, Hague T, Berces A, Juhos S. (2013) HLA typing from 1000 genomes whole genome and whole exome illumina data. PLoS One 8: e78410. 10.1371/journal.pone.0078410; 10.1371/journal.pone.0078410.
101. Lind C, Ferriola D, Mackiewicz K, Heron S, Rogers M, et al. (2010) Next-generation sequencing: The solution for high-resolution, unambiguous human leukocyte antigen typing. Hum Immunol 71: 1033-1042. 10.1016/j.humimm.2010.06.016; 10.1016/j.humimm.2010.06.016.
102. Warren RL, Choe G, Freeman DJ, Castellarin M, Munro S, et al. (2012) Derivation of HLA types from shotgun sequence datasets. Genome Med 4: 95. 10.1186/gm396.
103. Sanchez-Mazas A. (2001) African diversity from the HLA point of view: Influence of genetic drift, geography, linguistics, and natural selection. Hum Immunol 62: 937-948.
104. Fernandez Vina MA, Hollenbach JA, Lyke KE, Sztein MB, Maiers M, et al. (2012) Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and haplotypes in closed and open populations. Philos Trans R Soc Lond B Biol Sci 367: 820-829. 10.1098/rstb.2011.0320; 10.1098/rstb.2011.0320.

105. Middleton D, Gonzalez F, Fernandez-Vina M, Tiercy JM, Marsh SG, et al. (2009) A bioinformatics approach to ascertaining the rarity of HLA alleles. Tissue Antigens 74: 480-485. 10.1111/j.1399-0039.2009.01361.x; 10.1111/j.1399-0039.2009.01361.x.
106. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, et al. (2002) The structure of haplotype blocks in the human genome. Science 296: 2225-2229. 10.1126/science.1069424.
107. Thompson EA, Neel JV. (1997) Allelic disequilibrium and allele frequency distribution as a function of social and demographic history. Am J Hum Genet 60: 197-204.
108. Traherne JA, Horton R, Roberts AN, Miretti MM, Hurles ME, et al. (2006) Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. PLoS Genet 2: e9. 10.1371/journal.pgen.0020009.
109. Sanchez-Mazas A, Lemaitre JF, Currat M. (2012) Distinct evolutionary strategies of human leucocyte antigen loci in pathogen-rich environments. Philos Trans R Soc Lond B Biol Sci 367: 830-839. 10.1098/rstb.2011.0312; 10.1098/rstb.2011.0312.
110. Spurgin LG, Richardson DS. (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc Biol Sci 277: 979-988. 10.1098/rspb.2009.2084; 10.1098/rspb.2009.2084.
111. Hedrick PW. (2002) Pathogen resistance and genetic variation at MHC loci. Evolution 56: 1902-1908.
112. Riccio ME, Buhler S, Nunes JM, Vangenot C, Cuenod M, et al. (2013) 16(th) IHIW: Analysis of HLA population data, with updated results for 1996 to 2012 workshop data (AHPD project report). Int J Immunogenet 40: 21-30. 10.1111/iji.12033; 10.1111/iji.12033.
113. Kennedy LJ, Barnes A, Short A, Brown JJ, Seddon J, et al. (2007) Canine DLA diversity: 3. disease studies. Tissue Antigens 69 Suppl 1: 292-296. 10.1111/j.1399-0039.2006.00781.x.
114. Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. (2011) Allele frequency net: A database and online repository for immune gene frequencies in worldwide populations. Nucleic Acids Res 39: D913-9. 10.1093/nar/gkq1128; 10.1093/nar/gkq1128.
115. Sakaguchi S. (2005) Naturally arising Foxp3-expressing CD25+CD4+ regulatory T cells in immunological tolerance to self and non-self. Nat Immunol 6: 345-352. 10.1038/ni1178.
116. Bettens F, Passweg J, Schanz U, Chalandon Y, Heim D, et al. (2012) Impact of HLA-DPB1 haplotypes on outcome of 10/10 matched unrelated hematopoietic stem cell donor transplants depends on MHC-linked microsatellite polymorphisms. Biol Blood Marrow Transplant 18: 608-616. 10.1016/j.bbmt.2011.09.011; 10.1016/j.bbmt.2011.09.011.
117. Shaw BE, Arguello R, Garcia-Sepulveda CA, Madrigal JA. (2010) The impact of HLA genotyping on survival following unrelated donor haematopoietic stem cell transplantation. Br J Haematol 150: 251-258. 10.1111/j.1365-2141.2010.08224.x; 10.1111/j.1365-2141.2010.08224.x.
118. Johansen KA, Schneider JF, McCaffree MA, Woods GL, Council on Science and Public Health, American Medical Association. (2008) Efforts of the united states' national marrow donor program and registry to improve utilization and representation of minority donors. Transfus Med 18: 250-259. 10.1111/j.1365-3148.2008.00865.x; 10.1111/j.1365-3148.2008.00865.x.
119. Ottinger HD, Ferencik S, Beelen DW, Lindemann M, Peceny R, et al. (2003) Hematopoietic stem cell transplantation: Contrasting the outcome of transplantations from HLA-identical siblings, partially HLA-mismatched related donors, and HLA-matched unrelated donors. Blood 102: 1131-1137. 10.1182/blood-2002-09-2866.
120. Lee SJ, Klein J, Haagenson M, Baxter-Lowe LA, Confer DL, et al. (2007) High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. Blood 110: 4576-4583. 10.1182/blood-2007-06-097386.
121. Kawase T, Morishima Y, Matsuo K, Kashiwase K, Inoko H, et al. (2007) High-risk HLA allele mismatch combinations responsible for severe acute graft-versus-host disease and implication for its molecular mechanism. Blood 110: 2235-2241. 10.1182/blood-2007-02-072405.
122. Petersdorf EW, Malkki M, Horowitz MM, Spellman SR, Haagenson MD, et al. (2013) Mapping MHC haplotype effects in unrelated donor hematopoietic cell transplantation. Blood 121: 1896-1905. 10.1182/blood-2012-11-465161; 10.1182/blood-2012-11-465161.

123. Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678. 10.1038/nature05911.
124. Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, et al. (1989) Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. J Exp Med 169: 345-350.
125. Karell K, Louka AS, Moodie SJ, Ascher H, Clot F, et al. (2003) HLA types in celiac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: Results from the european genetics cluster on celiac disease. Hum Immunol 64: 469-477.
126. Louka AS, Moodie SJ, Karell K, Bolognesi E, Ascher H, et al. (2003) A collaborative european search for non-DQA1*05-DQB1*02 celiac disease loci on HLA-DR3 haplotypes: Analysis of transmission from homozygous parents. Hum Immunol 64: 350-358.
127. Simmonds MJ, Heward JM, Barrett JC, Franklyn JA, Gough SC. (2006) Association of the BTNL2 rs2076530 single nucleotide polymorphism with graves' disease appears to be secondary to DRB1 exon 2 position beta74. Clin Endocrinol (Oxf) 65: 429-432. 10.1111/j.1365-2265.2006.02586.x.
128. Tamai H, Tanaka K, Komaki G, Matsubayashi S, Hirota Y, et al. (1987) HLA and thyrotoxic periodic paralysis in japanese patients. J Clin Endocrinol Metab 64: 1075-1078.
129. Fong KY, Howe HS, Tin SK, Boey ML, Feng PH. (1996) Polymorphism of the regulatory region of tumour necrosis factor alpha gene in patients with systemic lupus erythematosus. Ann Acad Med Singapore 25: 90-93.
130. Singal DP, Blajchman MA. (1973) Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. Diabetes 22: 429-432.
131. Cudworth AG, Woodrow JC. (1975) Evidence for HL-A-linked genes in "juvenile" diabetes mellitus. Br Med J 3: 133-135.
132. Ilonen J, Herva E, Tiilikainen A, Akerblom HK, Koivukangas T, et al. (1978) HLA-Dw2 as a marker of resistance against juvenile diabetes mellitus. Tissue Antigens 11: 144-146.
133. Dorman JS, Bunker CH. (2000) HLA-DQ locus of the human leukocyte antigen complex and type 1 diabetes mellitus: A HuGE review. Epidemiol Rev 22: 218-227.
134. Lee YJ, Huang FY, Wang CH, Lo FS, Tsan KW, et al. (2000) Polymorphism in the transmembrane region of the MICA gene and type 1 diabetes. J Pediatr Endocrinol Metab 13: 489-496.
135. Hillert J, Olerup O. (1993) Multiple sclerosis is associated with genes within or close to the HLA-DR-DQ subregion on a normal DR15,DQ6,Dw2 haplotype. Neurology 43: 163-168.
136. Gao XJ, Brautbar C, Gazit E, Segal R, Naparstek Y, et al. (1991) A variant of HLA-DR4 determines susceptibility to rheumatoid arthritis in a subset of israeli jews. Arthritis Rheum 34: 547-551.
137. Milicic A, Lee D, Brown MA, Darke C, Wordsworth BP. (2002) HLA-DR/DQ haplotype in rheumatoid arthritis: Novel allelic associations in UK caucasians. J Rheumatol 29: 1821-1826.
138. Newton JL, Harney SM, Wordsworth BP, Brown MA. (2004) A review of the MHC genetics of rheumatoid arthritis. Genes Immun 5: 151-157. 10.1038/sj.gene.6364045.
139. de Vries N, Tijssen H, van Riel PL, van de Putte LB. (2002) Reshaping the shared epitope hypothesis: HLA-associated risk for rheumatoid arthritis is encoded by amino acid substitutions at positions 67-74 of the HLA-DRB1 molecule. Arthritis Rheum 46: 921-928.
140. Brinkman BM, Huizinga TW, Kurban SS, van der Velde EA, Schreuder GM, et al. (1997) Tumour necrosis factor alpha gene polymorphisms in rheumatoid arthritis: Association with susceptibility to, or severity of, disease? Br J Rheumatol 36: 516-521.
141. Brewerton DA, Hart FD, Nicholls A, Caffrey M, James DC, et al. (1973) Ankylosing spondylitis and HL-A 27. Lancet 1: 904-907.
142. Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, 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. 10.1086/518257.
143. Reveille JD, Moulds JM, Ahn C, Friedman AW, Baethge B, et al. (1998) Systemic lupus erythematosus in three ethnic groups: I. the effects of HLA class II, C4, and CR1 alleles,

socioeconomic factors, and ethnicity at disease onset. LUMINA study group. lupus in minority populations, nature versus nurture. Arthritis Rheum 41: 1161-1172. 2-K.
144. Mignot E, Lin X, Arrigoni J, Macaubas C, Olive F, et al. (1994) DQB1*0602 and DQA1*0102 (DQ1) are better markers than DR2 for narcolepsy in caucasian and black americans. Sleep 17: S60-7.
145. Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, et al. (2010) Common genetic variation in the HLA region is associated with late-onset sporadic parkinson's disease. Nat Genet 42: 781-785. 10.1038/ng.642; 10.1038/ng.642.
146. Illing PT, Vivian JP, Purcell AW, Rossjohn J, McCluskey J. (2013) Human leukocyte antigen-associated drug hypersensitivity. Curr Opin Immunol 25: 81-89. 10.1016/j.coi.2012.10.002; 10.1016/j.coi.2012.10.002.
147. Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, et al. (2008) Defining the role of the MHC in autoimmunity: A review and pooled analysis. PLoS Genet 4: e1000024. 10.1371/journal.pgen.1000024; 10.1371/journal.pgen.1000024.
148. Tait KF, Gough SC. (2003) The genetics of autoimmune endocrine disease. Clin Endocrinol (Oxf) 59: 1-11.
149. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, et al. (2008) Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med 359: 2767-2777. 10.1056/NEJMoa0807917; 10.1056/NEJMoa0807917.
150. Candore G, Modica MA, Lio D, Colonna-Romano G, Listi F, et al. (2003) Pathogenesis of autoimmune diseases associated with 8.1 ancestral haplotype: A genetically determined defect of C4 influences immunological parameters of healthy carriers of the haplotype. Biomed Pharmacother 57: 274-277.
151. Tabary T, Lehoang P, Betuel H, Benhamou A, Semiglia R, et al. (1990) Susceptibility to birdshot chorioretinopathy is restricted to the HLA-A29.2 subtype. Tissue Antigens 36: 177-179.
152. Wellcome Trust Case Control Consortium, Australo-Anglo-American Spondylitis Consortium (TASC), Burton PR, Clayton DG, Cardon LR, et al. (2007) Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 39: 1329-1337. 10.1038/ng.2007.17.
153. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, et al. (2011) Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet 43: 761-767. 10.1038/ng.873; 10.1038/ng.873.
154. Schon MP, Boehncke WH. (2005) Psoriasis. N Engl J Med 352: 1899-1912. 10.1056/NEJMra041320.
155. Palikhe A, Sinisalo J, Seppanen M, Valtonen V, Nieminen MS, et al. (2007) Human MHC region harbors both susceptibility and protective haplotypes for coronary artery disease. Tissue Antigens 69: 47-55. 10.1111/j.1399-0039.2006.00735.x.
156. Maclaren NK, Riley WJ. (1986) Inherited susceptibility to autoimmune addison's disease is linked to human leukocyte antigens-DR3 and/or DR4, except when associated with type I autoimmune polyglandular syndrome. J Clin Endocrinol Metab 62: 455-459.
157. Mackay IR. (2008) Historical reflections on autoimmune hepatitis. World J Gastroenterol 14: 3292-3300.
158. Chen QY, Huang W, She JX, Baxter F, Volpe R, et al. (1999) HLA-DRB1*08, DRB1*03/DRB3*0101, and DRB3*0202 are susceptibility genes for graves' disease in north american caucasians, whereas DRB1*07 is protective. J Clin Endocrinol Metab 84: 3182-3186.
159. Degli-Esposti MA, Abraham LJ, McCann V, Spies T, Christiansen FT, et al. (1992) Ancestral haplotypes reveal the role of the central MHC in the immunogenetics of IDDM. Immunogenetics 36: 345-356.
160. Hietaharju A, Korpela M, Ilonen J, Frey H. (1992) Nervous system disease, immunological features, and HLA phenotype in sjogren's syndrome. Ann Rheum Dis 51: 506-509.
161. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, et al. (2008) Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nat Genet 40: 1059-1061. 10.1038/ng.200; 10.1038/ng.200.
162. Hedfors E, Lindstrom F. (1983) HLA-B8/DR3 in sarcoidosis. correlation to acute onset disease with arthritis. Tissue Antigens 22: 200-203.

163. Berlin M, Fogdell-Hahn A, Olerup O, Eklund A, Grunewald J. (1997) HLA-DR predicts the prognosis in scandinavian patients with pulmonary sarcoidosis. Am J Respir Crit Care Med 156: 1601-1605.
164. Brewerton DA, Cockburn C, James DC, James DG, Neville E. (1977) HLA antigens in sarcoidosis. Clin Exp Immunol 27: 227-229.
165. Spurkland A, Sollid LM, Polanco I, Vartdal F, Thorsby E. (1992) HLA-DR and -DQ genotypes of celiac disease patients serologically typed to be non-DR3 or non-DR5/7. Hum Immunol 35: 188-192.
166. Jones EY, Fugger L, Strominger JL, Siebold C. (2006) MHC class II proteins and disease: A structural perspective. Nat Rev Immunol 6: 271-282. 10.1038/nri1805.
167. Simmonds MJ, Gough SC. (2005) Genetic insights into disease mechanisms of autoimmunity. Br Med Bull 71: 93-113. 10.1093/bmb/ldh032.
168. Traherne JA. (2008) Human MHC architecture and evolution: Implications for disease association studies. Int J Immunogenet 35: 179-192. 10.1111/j.1744-313X.2008.00765.x; 10.1111/j.1744-313X.2008.00765.x.
169. Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, et al. (2008) HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: Analysis of the type 1 diabetes genetics consortium families. Diabetes 57: 1084-1092. 10.2337/db07-1331; 10.2337/db07-1331.
170. Australo-Anglo-American Spondyloarthritis Consortium (TASC), Reveille JD, Sims AM, Danoy P, Evans DM, et al. (2010) Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nat Genet 42: 123-127. 10.1038/ng.513; 10.1038/ng.513.
171. Lindfors K, Maki M, Kaukinen K. (2010) Transglutaminase 2-targeted autoantibodies in celiac disease: Pathogenetic players in addition to diagnostic tools? Autoimmun Rev 9: 744-749. 10.1016/j.autrev.2010.06.003; 10.1016/j.autrev.2010.06.003.
172. Medica I, Kastrin A, Maver A, Peterlin B. (2007) Role of genetic polymorphisms in ACE and TNF-alpha gene in sarcoidosis: A meta-analysis. J Hum Genet 52: 836-847. 10.1007/s10038-007-0185-7.
173. Steenvoorden MM, Toes RE, Ronday HK, Huizinga TW, Degroot J. (2007) RAGE activation induces invasiveness of RA fibroblast-like synoviocytes in vitro. Clin Exp Rheumatol 25: 740-742.
174. Campo I, Morbini P, Zorzetto M, Tinelli C, Brunetta E, et al. (2007) Expression of receptor for advanced glycation end products in sarcoid granulomas. Am J Respir Crit Care Med 175: 498-506. 10.1164/rccm.200601-136OC.
175. Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, et al. (2005) Sarcoidosis is associated with a truncating splice site mutation in BTNL2. Nat Genet 37: 357-364. 10.1038/ng1519.
176. Rybicki BA, Iannuzzi MC. (2007) Epidemiology of sarcoidosis: Recent advances and future prospects. Semin Respir Crit Care Med 28: 22-35. 10.1055/s-2007-970331.
177. Newman LS, Rose CS, Bresnitz EA, Rossman MD, Barnard J, et al. (2004) A case control etiologic study of sarcoidosis: Environmental and occupational risk factors. Am J Respir Crit Care Med 170: 1324-1330. 10.1164/rccm.200402-249OC.
178. Sverrild A, Backer V, Kyvik KO, Kaprio J, Milman N, et al. (2008) Heredity in sarcoidosis: A registry-based twin study. Thorax 63: 894-896. 10.1136/thx.2007.094060.
179. Rybicki BA, Kirkey KL, Major M, Maliarik MJ, Popovich J,Jr, et al. (2001) Familial risk ratio of sarcoidosis in african-american sibs and parents. Am J Epidemiol 153: 188-193.
180. Statement on sarcoidosis. (1999) Statement on sarcoidosis. joint statement of the american thoracic society (ATS), the european respiratory society (ERS) and the world association of sarcoidosis and other granulomatous disorders (WASOG) adopted by the ATS board of directors and by the ERS executive committee, february 1999. Am J Respir Crit Care Med 160: 736-755.
181. Rybicki BA, Maliarik MJ, Major M, Popovich J,Jr, Iannuzzi MC. (1998) Epidemiology, demographics, and genetics of sarcoidosis. Semin Respir Infect 13: 166-173.
182. Rybicki BA, Major M, Popovich J,Jr, Maliarik MJ, Iannuzzi MC. (1997) Racial differences in sarcoidosis incidence: A 5-year study in a health maintenance organization. Am J Epidemiol 145: 234-241.
183. Morimoto T, Azuma A, Abe S, Usuki J, Kudoh S, et al. (2008) Epidemiology of sarcoidosis in japan. Eur Respir J 31: 372-379. 10.1183/09031936.00075307.

184. Milman N, Selroos O. (1990) Pulmonary sarcoidosis in the nordic countries 1950-1982. epidemiology and clinical picture. Sarcoidosis 7: 50-57.
185. Lynch JP,3rd, Ma YL, Koss MN, White ES. (2007) Pulmonary sarcoidosis. Semin Respir Crit Care Med 28: 53-74. 10.1055/s-2007-970333.
186. Judson MA. (2007) Extrapulmonary sarcoidosis. Semin Respir Crit Care Med 28: 83-101. 10.1055/s-2007-970335.
187. Valeyre D, Prasse A, Nunes H, Uzunhan Y, Brillet PY, et al. (2013) Sarcoidosis. Lancet . 10.1016/S0140-6736(13)60680-7; 10.1016/S0140-6736(13)60680-7.
188. Pietinalho A, Ohmichi M, Hiraga Y, Lofroos AB, Selroos O. (1996) The mode of presentation of sarcoidosis in finland and hokkaido, japan. A comparative analysis of 571 finnish and 686 japanese patients. Sarcoidosis Vasc Diffuse Lung Dis 13: 159-166.
189. Lagana SM, Parwani AV, Nichols LC. (2010) Cardiac sarcoidosis: A pathology-focused review. Arch Pathol Lab Med 134: 1039-1046. 10.1043/2009-0274-RA.1; 10.1043/2009-0274-RA.1.
190. Nunes H, Bouvry D, Soler P, Valeyre D. (2007) Sarcoidosis. Orphanet J Rare Dis 2: 46. 10.1186/1750-1172-2-46.
191. Zissel G, Prasse A, Muller-Quernheim J. (2010) Immunologic response of sarcoidosis. Semin Respir Crit Care Med 31: 390-403. 10.1055/s-0030-1262208; 10.1055/s-0030-1262208.
192. Grunewald J, Eklund A. (2009) Lofgren's syndrome: Human leukocyte antigen strongly influences the disease course. Am J Respir Crit Care Med 179: 307-312. 10.1164/rccm.200807-1082OC.
193. Baughman RP, Drent M, Kavuru M, Judson MA, Costabel U, et al. (2006) Infliximab therapy in patients with chronic sarcoidosis and pulmonary involvement. Am J Respir Crit Care Med 174: 795-802. 10.1164/rccm.200603-402OC.
194. Judson MA, Baughman RP, Costabel U, Flavin S, Lo KH, et al. (2008) Efficacy of infliximab in extrapulmonary sarcoidosis: Results from a randomised trial. Eur Respir J 31: 1189-1196. 10.1183/09031936.00051907; 10.1183/09031936.00051907.
195. Bradley B, Branley HM, Egan JJ, Greaves MS, Hansell DM, et al. (2008) Interstitial lung disease guideline: The british thoracic society in collaboration with the thoracic society of australia and new zealand and the irish thoracic society. Thorax 63 Suppl 5: v1-58. 10.1136/thx.2008.101691; 10.1136/thx.2008.101691.
196. Wiken M, Idali F, Al Hayja MA, Grunewald J, Eklund A, et al. (2010) No evidence of altered alveolar macrophage polarization, but reduced expression of TLR2, in bronchoalveolar lavage cells in sarcoidosis. Respir Res 11: 121-9921-11-121. 10.1186/1465-9921-11-121; 10.1186/1465-9921-11-121.
197. Oswald-Richter KA, Culver DA, Hawkins C, Hajizadeh R, Abraham S, et al. (2009) Cellular responses to mycobacterial antigens are present in bronchoalveolar lavage fluid used in the diagnosis of sarcoidosis. Infect Immun 77: 3740-3748. 10.1128/IAI.00142-09; 10.1128/IAI.00142-09.
198. Kucera GP, Rybicki BA, Kirkey KL, Coon SW, Major ML, et al. (2003) Occupational risk factors for sarcoidosis in african-american siblings. Chest 123: 1527-1535.
199. Rybicki BA, Maliarik MJ, Poisson LM, Iannuzzi MC. (2004) Sarcoidosis and granuloma genes: A family-based study in african-americans. Eur Respir J 24: 251-257.
200. Rossman MD, Thompson B, Frederick M, Iannuzzi MC, Rybicki BA, et al. (2008) HLA and environmental interactions in sarcoidosis. Sarcoidosis Vasc Diffuse Lung Dis 25: 125-132.
201. Deubelbeiss U, Gemperli A, Schindler C, Baty F, Brutsche MH. (2010) Prevalence of sarcoidosis in switzerland is associated with environmental factors. Eur Respir J 35: 1088-1097. 10.1183/09031936.00197808; 10.1183/09031936.00197808.
202. Almadi MA, Aljebreen AM, Sanai FM, Marcus V, Almeghaiseeb ES, et al. (2011) New insights into gastrointestinal and hepatic granulomatous disorders. Nat Rev Gastroenterol Hepatol 8: 455-466. 10.1038/nrgastro.2011.115; 10.1038/nrgastro.2011.115.
203. Soler P, Basset F. (1976) Morphology and distribution of the cells of a sarcoid granuloma: Ultrastructural study of serial sections. Ann N Y Acad Sci 278: 147-160.
204. Agostini C, Adami F, Semenzato G. (2000) New pathogenetic insights into the sarcoid granuloma. Curr Opin Rheumatol 12: 71-76.
205. Semenzato G, Pezzutto A, Chilosi M, Pizzolo G. (1982) Redistribution of T lymphocytes in the lymph nodes of patients with sarcoidosis. N Engl J Med 306: 48-49. 10.1056/NEJM198201073060114.

206. Greene CM, Meachery G, Taggart CC, Rooney CP, Coakley R, et al. (2000) Role of IL-18 in CD4+ T lymphocyte activation in sarcoidosis. J Immunol 165: 4718-4724.
207. Welker L, Jorres RA, Costabel U, Magnussen H. (2004) Predictive value of BAL cell differentials in the diagnosis of interstitial lung diseases. Eur Respir J 24: 1000-1006. 10.1183/09031936.04.00101303.
208. Miyara M, Amoura Z, Parizot C, Badoual C, Dorgham K, et al. (2006) The immune paradox of sarcoidosis and regulatory T cells. J Exp Med 203: 359-370. 10.1084/jem.20050648.
209. Idali F, Wahlstrom J, Dahlberg B, Khademi M, Olsson T, et al. (2009) Altered expression of T cell immunoglobulin-mucin (TIM) molecules in bronchoalveolar lavage CD4+ T cells in sarcoidosis. Respir Res 10: 42. 10.1186/1465-9921-10-42.
210. Rybicki BA, Iannuzzi MC, Frederick MM, Thompson BW, Rossman MD, et al. (2001) Familial aggregation of sarcoidosis. A case-control etiologic study of sarcoidosis (ACCESS). Am J Respir Crit Care Med 164: 2085-2091.
211. Sato H, Woodhead FA, Ahmad T, Grutters JC, Spagnolo P, et al. (2010) Sarcoidosis HLA class II genotyping distinguishes differences of clinical phenotype across ethnic groups. Hum Mol Genet 19: 4100-4111. 10.1093/hmg/ddq325.
212. Spagnolo P, Grunewald J. (2013) Recent advances in the genetics of sarcoidosis. J Med Genet 50: 290-297. 10.1136/jmedgenet-2013-101532; 10.1136/jmedgenet-2013-101532.
213. Grunewald J, Eklund A, Olerup O. (2004) Human leukocyte antigen class I alleles and the disease course in sarcoidosis patients. Am J Respir Crit Care Med 169: 696-702. 10.1164/rccm.200303-459OC.
214. Rossman MD, Thompson B, Frederick M, Maliarik M, Iannuzzi MC, et al. (2003) HLA-DRB1*1101: A significant risk factor for sarcoidosis in blacks and whites. Am J Hum Genet 73: 720-735. 10.1086/378097.
215. Voorter CE, Amicosante M, Berretta F, Groeneveld L, Drent M, et al. (2007) HLA class II amino acid epitopes as susceptibility markers of sarcoidosis. Tissue Antigens 70: 18-27. 10.1111/j.1399-0039.2007.00842.x.
216. Wijnen PA, Nelemans PJ, Verschakelen JA, Bekers O, Voorter CE, et al. (2010) The role of tumor necrosis factor alpha G-308A polymorphisms in the course of pulmonary sarcoidosis. Tissue Antigens 75: 262-268. 10.1111/j.1399-0039.2009.01437.x.
217. Maier LA, McGrath DS, Sato H, Lympany P, Welsh K, et al. (2003) Influence of MHC class II in susceptibility to beryllium sensitization and chronic beryllium disease. J Immunol 171: 6910-6918.
218. McCanlies EC, Kreiss K, Andrew M, Weston A. (2003) HLA-DPB1 and chronic beryllium disease: A HuGE review. Am J Epidemiol 157: 388-398.
219. Richeldi L, Sorrentino R, Saltini C. (1993) HLA-DPB1 glutamate 69: A genetic marker of beryllium disease. Science 262: 242-244.
220. Sato H, Spagnolo P, Silveira L, Welsh KI, du Bois RM, et al. (2007) BTNL2 allele associations with chronic beryllium disease in HLA-DPB1*Glu69-negative individuals. Tissue Antigens 70: 480-486. 10.1111/j.1399-0039.2007.00944.x.
221. McIntyre JA, McKee KT, Loadholt CB, Mercurio S, Lin I. (1977) Increased HLA-B7 antigen frequency in south carolina blacks in association with sarcoidosis. Transplant Proc 9: 173-176.
222. Gardner J, Kennedy HG, Hamblin A, Jones E. (1984) HLA associations in sarcoidosis: A study of two ethnic groups. Thorax 39: 19-22.
223. Martinetti M, Tinelli C, Kolek V, Cuccia M, Salvaneschi L, et al. (1995) "The sarcoidosis map": A joint survey of clinical and immunogenetic findings in two european countries. Am J Respir Crit Care Med 152: 557-564. 10.1164/ajrccm.152.2.7633707.
224. Grunewald J, Brynedal B, Darlington P, Nisell M, Cederlund K, et al. (2010) Different HLA-DRB1 allele distributions in distinct clinical subgroups of sarcoidosis patients. Respir Res 11: 25. 10.1186/1465-9921-11-25.
225. Darlington P, Tallstedt L, Padyukov L, Kockum I, Cederlund K, et al. (2011) HLA-DRB1* alleles and symptoms associated with heerfordt's syndrome in sarcoidosis. Eur Respir J 38: 1151-1157. 10.1183/09031936.00025011.
226. Suzuki H, Ota M, Meguro A, Katsuyama Y, Kawagoe T, et al. (2012) Genetic characterization and susceptibility for sarcoidosis in japanese patients: Risk factors of BTNL2 gene polymorphisms and HLA class II alleles. Invest Ophthalmol Vis Sci 53: 7109-7115. 10.1167/iovs.12-10491; 10.1167/iovs.12-10491.

227. Iannuzzi MC, Maliarik MJ, Poisson LM, Rybicki BA. (2003) Sarcoidosis susceptibility and resistance HLA-DQB1 alleles in african americans. Am J Respir Crit Care Med 167: 1225-1231. 10.1164/rccm.200209-1097OC.
228. Voorter CE, Drent M, Hoitsma E, Faber KG, van den Berg-Loonen EM. (2005) Association of HLA DQB1 0602 in sarcoidosis patients with small fiber neuropathy. Sarcoidosis Vasc Diffuse Lung Dis 22: 129-132.
229. Voorter CE, Drent M, van den Berg-Loonen EM. (2005) Severe pulmonary sarcoidosis is strongly associated with the haplotype HLA-DQB1*0602-DRB1*150101. Hum Immunol 66: 826-835. 10.1016/j.humimm.2005.04.003.
230. Rybicki BA, Walewski JL, Maliarik MJ, Kian H, Iannuzzi MC, et al. (2005) The BTNL2 gene and sarcoidosis susceptibility in african americans and whites. Am J Hum Genet 77: 491-499. 10.1086/444435.
231. Strausz J, Mannel DN, Pfeifer S, Borkowski A, Ferlinz R, et al. (1991) Spontaneous monokine release by alveolar macrophages in chronic sarcoidosis. Int Arch Allergy Appl Immunol 96: 68-75.
232. Nguyen T, Liu XK, Zhang Y, Dong C. (2006) BTNL2, a butyrophilin-like molecule that functions to inhibit T cell activation. J Immunol 176: 7354-7360.
233. Spagnolo P, Sato H, Grutters JC, Renzoni EA, Marshall SE, et al. (2007) Analysis of BTNL2 genetic polymorphisms in british and dutch patients with sarcoidosis. Tissue Antigens 70: 219-227. 10.1111/j.1399-0039.2007.00879.x.
234. Wijnen PA, Voorter CE, Nelemans PJ, Verschakelen JA, Bekers O, et al. (2011) Butyrophilin-like 2 in pulmonary sarcoidosis: A factor for susceptibility and progression? Hum Immunol 72: 342-347. 10.1016/j.humimm.2011.01.011; 10.1016/j.humimm.2011.01.011.
235. Orozco G, Eerligh P, Sanchez E, Zhernakova S, Roep BO, et al. (2005) Analysis of a functional BTNL2 polymorphism in type 1 diabetes, rheumatoid arthritis, and systemic lupus erythematosus. Hum Immunol 66: 1235-1241. 10.1016/j.humimm.2006.02.003.
236. Johnson CM, Traherne JA, Jamieson SE, Tremelling M, Bingham S, et al. (2007) Analysis of the BTNL2 truncating splice site mutation in tuberculosis, leprosy and crohn's disease. Tissue Antigens 69: 236-241. 10.1111/j.1399-0039.2006.00795.x.
237. Adrianto I, Lin CP, Hale JJ, Levin AM, Datta I, et al. (2012) Genome-wide association study of african and european americans implicates multiple shared and ethnic specific loci in sarcoidosis susceptibility. PLoS One 7: e43907. 10.1371/journal.pone.0043907; 10.1371/journal.pone.0043907.
238. Kruit A, Ruven HJ, Grutters JC, van den Bosch JM. (2010) Angiotensin II receptor type 1 1166 A/C and angiotensin converting enzyme I/D gene polymorphisms in a dutch sarcoidosis cohort. Sarcoidosis Vasc Diffuse Lung Dis 27: 147-152.
239. Rybicki BA, Iannuzzi MC. (2004) Sarcoidosis and human leukocyte antigen class I and II genes: It takes two to tango? Am J Respir Crit Care Med 169: 665-666. 10.1164/rccm.2401005.
240. Fischer A, Schmid B, Ellinghaus D, Nothnagel M, Gaede KI, et al. (2012) A novel sarcoidosis risk locus for europeans on chromosome 11q13.1. Am J Respir Crit Care Med 186: 877-885. 10.1164/rccm.201204-0708OC; 10.1164/rccm.201204-0708OC.
241. Grunewald J. (2010) Review: Role of genetics in susceptibility and outcome of sarcoidosis. Semin Respir Crit Care Med 31: 380-389. 10.1055/s-0030-1262206.
242. Hofmann S, Franke A, Fischer A, Jacobs G, Nothnagel M, et al. (2008) Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. Nat Genet 40: 1103-1106. 10.1038/ng.198.
243. Buschow SI, van Balkom BW, Aalberts M, Heck AJ, Wauben M, et al. (2010) MHC class II-associated proteins in B-cell exosomes and potential functional implications for exosome biogenesis. Immunol Cell Biol 88: 851-856. 10.1038/icb.2010.64; 10.1038/icb.2010.64.
244. Levin AM, Iannuzzi MC, Montgomery CG, Trudeau S, Datta I, et al. (2013) Association of ANXA11 genetic variation with sarcoidosis in african americans and european americans. Genes Immun 14: 13-18. 10.1038/gene.2012.48; 10.1038/gene.2012.48.
245. Mrazek F, Stahelova A, Kriegova E, Fillerova R, Zurkova M, et al. (2011) Functional variant ANXA11 R230C: True marker of protection and candidate disease modifier in sarcoidosis. Genes Immun 12: 490-494. 10.1038/gene.2011.27; 10.1038/gene.2011.27.

246. Li Y, Pabst S, Kubisch C, Grohe C, Wollnik B. (2010) First independent replication study confirms the strong genetic association of ANXA11 with sarcoidosis. Thorax 65: 939-940. 10.1136/thx.2010.138743; 10.1136/thx.2010.138743.
247. Franke A, Fischer A, Nothnagel M, Becker C, Grabe N, et al. (2008) Genome-wide association analysis in sarcoidosis and crohn's disease unravels a common susceptibility locus on 10p12.2. Gastroenterology 135: 1207-1215. 10.1053/j.gastro.2008.07.017.
248. Schurmann M, Reichel P, Muller-Myhsok B, Schlaak M, Muller-Quernheim J, et al. (2001) Results from a genome-wide search for predisposing genes in sarcoidosis. Am J Respir Crit Care Med 164: 840-846.
249. Morohashi K, Takada T, Omori K, Suzuki E, Gejyo F. (2003) Vascular endothelial growth factor gene polymorphisms in japanese patients with sarcoidosis. Chest 123: 1520-1526.
250. Hofmann S, Fischer A, Nothnagel M, Jacobs G, Schmid B, et al. (2013) Genome-wide association analysis reveals 12q13.3-q14.1 as new risk locus for sarcoidosis. Eur Respir J 41: 888-900. 10.1183/09031936.00033812; 10.1183/09031936.00033812.
251. Fischer A, Nothnagel M, Franke A, Jacobs G, Saadati HR, et al. (2010) Association of IBD risk loci with sarcoidosis and its acute and chronic subphenotypes. Eur Respir J . 10.1183/09031936.00049410.
252. Seppanen M, Suvilehto J, Lokki ML, Notkola IL, Jarvinen A, et al. (2006) Immunoglobulins and complement factor C4 in adult rhinosinusitis. Clin Exp Immunol 145: 219-227. 10.1111/j.1365-2249.2006.03134.x.
253. Pietinalho A, Tukiainen P, Haahtela T, Persson T, Selroos O, et al. (2002) Early treatment of stage II sarcoidosis improves 5-year pulmonary function. Chest 121: 24-31.
254. Heron M, van Moorsel CH, Grutters JC, Huizinga TW, van der Helm-van Mil AH, et al. (2011) Genetic variation in GREM1 is a risk factor for fibrosis in pulmonary sarcoidosis. Tissue Antigens 77: 112-117. 10.1111/j.1399-0039.2010.01590.x; 10.1111/j.1399-0039.2010.01590.x.
255. Foley PJ, McGrath DS, Puscinska E, Petrek M, Kolek V, et al. (2001) Human leukocyte antigen-DRB1 position 11 residues are a common protective marker for sarcoidosis. Am J Respir Cell Mol Biol 25: 272-277.
256. Veltkamp M, van Moorsel CH, Rijkers GT, Ruven HJ, Grutters JC. (2012) Genetic variation in the toll-like receptor gene cluster (TLR10-TLR1-TLR6) influences disease course in sarcoidosis. Tissue Antigens 79: 25-32. 10.1111/j.1399-0039.2011.01808.x; 10.1111/j.1399-0039.2011.01808.x.
257. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575. 10.1086/519795.
258. Paakkanen R, Vauhkonen H, Eronen KT, Jarvinen A, Seppanen M, et al. (2012) Copy number analysis of complement C4A, C4B and C4A silencing mutation by real-time quantitative polymerase chain reaction. PLoS One 7: e38813. 10.1371/journal.pone.0038813; 10.1371/journal.pone.0038813.
259. Hollenbach JA, Madbouly A, Gragert L, Vierra-Green C, Flesch S, et al. (2012) A combined DPA1~DPB1 amino acid epitope is the primary unit of selection on the HLA-DP heterodimer. Immunogenetics 64: 559-569. 10.1007/s00251-012-0615-3.
260. Excoffier L, Laval G, Schneider S. (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol Bioinform Online 1: 47-50.
261. Rousset F. (2008) Genepop'007: A complete re-implementation of the genepop software for windows and linux. Mol Ecol Resour 8: 103-106. 10.1111/j.1471-8286.2007.01931.x; 10.1111/j.1471-8286.2007.01931.x.
262. Barrett JC, Fry B, Maller J, Daly MJ. (2005) Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263-265. 10.1093/bioinformatics/bth457.
263. Stephens M, Smith NJ, Donnelly P. (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68: 978-989. 10.1086/319501.
264. Mack SJ, Tu B, Lazaro A, Yang R, Lancaster AK, et al. (2009) HLA-A, -B, -C, and -DRB1 allele and haplotype frequencies distinguish eastern european americans from the general european american population. Tissue Antigens 73: 17-32. 10.1111/j.1399-0039.2008.01151.x; 10.1111/j.1399-0039.2008.01151.x.
265. Hollenbach JA, Mack SJ, Gourraud PA, Single RM, Maiers M, et al. (2011) A community standard for immunogenomic data reporting and analysis: Proposal for a STrengthening

the REporting of immunogenomic studies statement. Tissue Antigens 78: 333-344. 10.1111/j.1399-0039.2011.01777.x; 10.1111/j.1399-0039.2011.01777.x.
266. Saitou N, Nei M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425.
267. Kanterakis S, Magira E, Rosenman KD, Rossman M, Talsania K, et al. (2008) SKDM human leukocyte antigen (HLA) tool: A comprehensive HLA and disease associations analysis software. Hum Immunol 69: 522-525. 10.1016/j.humimm.2008.05.011.
268. Williams F, Meenagh A, Maxwell AP, Middleton D. (1999) Allele resolution of HLA-A using oligonucleotide probes in a two-stage typing strategy. Tissue Antigens 54: 59-68.
269. Middleton D, Williams F, Hamill MA, Meenagh A. (2000) Frequency of HLA-B alleles in a caucasoid population determined by a two-stage PCR-SSOP typing strategy. Hum Immunol 61: 1285-1297.
270. Johansson A, Ingman M, Mack SJ, Erlich H, Gyllensten U. (2008) Genetic origin of the swedish sami inferred from HLA class I and class II allele frequencies. Eur J Hum Genet 16: 1341-1349. 10.1038/ejhg.2008.88; 10.1038/ejhg.2008.88.
271. Arnaiz-Villena A, Martinez-Laso J, Moscoso J, Livshits G, Zamora J, et al. (2003) HLA genes in the chuvashian population from european russia: Admixture of central european and mediterranean populations. Hum Biol 75: 375-392.
272. Lam TH, Shen M, Chia JM, Chan SH, Ren EC. (2013) Population-specific recombination sites within the human MHC region. Heredity (Edinb) . 10.1038/hdy.2013.27; 10.1038/hdy.2013.27.
273. Jeffreys AJ, Kauppi L, Neumann R. (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 29: 217-222. 10.1038/ng1001-217.
274. Kauppi L, Sajantila A, Jeffreys AJ. (2003) Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. Hum Mol Genet 12: 33-40.
275. Yunis EJ, Larsen CE, Fernandez-Vina M, Awdeh ZL, Romero T, et al. (2003) Inheritable variable sizes of DNA stretches in the human MHC: Conserved extended haplotypes and their fragments or blocks. Tissue Antigens 62: 1-20.
276. Furst D, Zollikofer C, Schrezenmeier H, Mytilineos J. (2012) TNFA promoter alleles--frequencies and linkage with classical HLA genes in a south german caucasian population. Tissue Antigens 80: 502-508. 10.1111/tan.12025; 10.1111/tan.12025.
277. Kainulainen L, Peltola V, Seppanen M, Viander M, He Q, et al. (2012) C4A deficiency in children and adolescents with recurrent respiratory infections. Hum Immunol 73: 498-501. 10.1016/j.humimm.2012.02.015; 10.1016/j.humimm.2012.02.015.
278. Sane J, Kurkela S, Desdouits M, Kalimo H, Mazalrey S, et al. (2012) Prolonged myalgia in sindbis virus infection: Case description and in vitro infection of myotubes and myoblasts. J Infect Dis 206: 407-414. 10.1093/infdis/jis358.
279. Seppanen M, Lokki ML, Notkola IL, Mattila K, Valtonen V, et al. (2007) Complement and c4 null alleles in severe chronic adult periodontitis. Scand J Immunol 65: 176-181. 10.1111/j.1365-3083.2006.01886.x.
280. Knight JC, Keating BJ, Kwiatkowski DP. (2004) Allele-specific repression of lymphotoxin-alpha by activated B cell factor-1. Nat Genet 36: 394-399. 10.1038/ng1331.
281. Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, et al. (2002) Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 32: 650-654. 10.1038/ng1047.
282. Benesova Y, Vasku A, Stourac P, Hladikova M, Fiala A, et al. (2013) Association of HLA-DRB1*1501 tagging rs3135388 gene polymorphism with multiple sclerosis. J Neuroimmunol 255: 92-96. 10.1016/j.jneuroim.2012.10.014; 10.1016/j.jneuroim.2012.10.014.
283. Samano ES, Ribeiro Lde M, Gorescu RG, Rocha KC, Grumach AS. (2004) Involvement of C4 allotypes in the pathogenesis of human diseases. Rev Hosp Clin Fac Med Sao Paulo 59: 138-144. /S0041-87812004000300009.
284. Jaatinen T, Lahti M, Ruuskanen O, Kinos R, Truedsson L, et al. (2003) Total C4B deficiency due to gene deletion and gene conversion in a patient with severe infections. Clin Diagn Lab Immunol 10: 195-201.

285. Lie BA, Thorsby E. (2005) Several genes in the extended human MHC contribute to predisposition to autoimmune diseases. Curr Opin Immunol 17: 526-531. 10.1016/j.coi.2005.07.001.
286. Jakkula E, Rehnstrom K, Varilo T, Pietilainen OP, Paunio T, et al. (2008) The genome-wide patterns of variation expose significant substructure in a founder population. Am J Hum Genet 83: 787-794. 10.1016/j.ajhg.2008.11.005; 10.1016/j.ajhg.2008.11.005.
287. Nelis M, Esko T, Magi R, Zimprich F, Zimprich A, et al. (2009) Genetic structure of europeans: A view from the north-east. PLoS One 4: e5472. 10.1371/journal.pone.0005472; 10.1371/journal.pone.0005472.
288. Siren MK, Sareneva H, Lokki ML, Koskimies S. (1996) Unique HLA antigen frequencies in the finnish population. Tissue Antigens 48: 703-707.
289. Buhler S, Nunes JM, Nicoloso G, Tiercy JM, Sanchez-Mazas A. (2012) The heterogeneous HLA genetic makeup of the swiss population. PLoS One 7: e41400. 10.1371/journal.pone.0041400; 10.1371/journal.pone.0041400.
290. Haimila K, Penttila A, Arvola A, Auvinen MK, Korhonen M. (2013) Analysis of the adequate size of a cord blood bank and comparison of HLA haplotype distributions between four populations. Hum Immunol 74: 189-195. 10.1016/j.humimm.2012.10.018; 10.1016/j.humimm.2012.10.018.
291. Haimila K, Perasaari J, Linjama T, Koskela S, Saarinenl T, et al. (2013) HLA antigen, allele and haplotype frequencies and their use in virtual panel reactive antigen calculations in the finnish population. Tissue Antigens 81: 35-43. 10.1111/tan.12036; 10.1111/tan.12036.
292. Rosenberg NA, Li LM, Ward R, Pritchard JK. (2003) Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 73: 1402-1422. 10.1086/380416.
293. Mack SJ, Tu B, Yang R, Masaberg C, Ng J, et al. (2011) Human leukocyte antigen-A, -B, -C, -DRB1 allele and haplotype frequencies in americans originating from southern europe: Contrasting patterns of population differentiation between italian and spanish americans. Hum Immunol 72: 144-149. 10.1016/j.humimm.2010.10.017; 10.1016/j.humimm.2010.10.017.
294. Hurley CK, Wagner JE, Setterholm MI, Confer DL. (2006) Advances in HLA: Practical implications for selecting adult donors and cord blood units. Biol Blood Marrow Transplant 12: 28-33. 10.1016/j.bbmt.2005.10.005.
295. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, et al. (2007) Convergent adaptation of human lactase persistence in africa and europe. Nat Genet 39: 31-40. 10.1038/ng1946.
296. McEvoy BP, Montgomery GW, McRae AF, Ripatti S, Perola M, et al. (2009) Geographical structure and differential natural selection among north european populations. Genome Res 19: 804-814. 10.1101/gr.083394.108; 10.1101/gr.083394.108.
297. Spagnolo P, Sato H, Grunewald J, Brynedal B, Hillert J, et al. (2008) A common haplotype of the C-C chemokine receptor 2 gene and HLA-DRB1*0301 are independent genetic risk factors for lofgren's syndrome. J Intern Med 264: 433-441. 10.1111/j.1365-2796.2008.01984.x; 10.1111/j.1365-2796.2008.01984.x.
298. Bogunia-Kubik K, Koscinska K, Suchnicki K, Lange A. (2006) HSP70-hom gene single nucleotide (+2763 G/A and +2437 C/T) polymorphisms in sarcoidosis. Int J Immunogenet 33: 135-140. 10.1111/j.1744-313X.2006.00584.x.
299. Morais A, Alves H, Lima B, Delgado L, Goncalves R, et al. (2008) HLA class I and II and TNF-alpha gene polymorphisms in sarcoidosis patients. Rev Port Pneumol 14: 727-746.
300. Bogunia-Kubik K, Tomeczko J, Suchnicki K, Lange A. (2001) HLA-DRB1*03, DRB1*11 or DRB1*12 and their respective DRB3 specificities in clinical variants of sarcoidosis. Tissue Antigens 57: 87-90.
301. Traherne JA, Barcellos LF, Sawcer SJ, Compston A, Ramsay PP, et al. (2006) Association of the truncating splice site mutation in BTNL2 with multiple sclerosis is secondary to HLA-DRB1*15. Hum Mol Genet 15: 155-161. 10.1093/hmg/ddi436.
302. Sun W, Cui Y, Zhen L, Huang L. (2011) Association between HLA-DRB1, HLA-DRQB1 alleles, and CD4(+)CD28(null) T cells in a chinese population with coronary heart disease. Mol Biol Rep 38: 1675-1679. 10.1007/s11033-010-0279-8.

303. Pazar B, Gergely P,Jr, Nagy ZB, Gombos T, Pozsonyi E, et al. (2008) Role of HLA-DRB1 and PTPN22 genes in susceptibility to juvenile idiopathic arthritis in hungarian patients. Clin Exp Rheumatol 26: 1146-1152.
304. Noble JA, Valdes AM, Thomson G, Erlich HA. (2000) The HLA class II locus DPB1 can influence susceptibility to type 1 diabetes. Diabetes 49: 121-125.
305. Lympany PA, Petrek M, Southcott AM, Newman Taylor AJ, Welsh KI, et al. (1996) HLA-DPB polymorphisms: Glu 69 association with sarcoidosis. Eur J Immunogenet 23: 353-359.
306. Milman N, Svendsen CB, Nielsen FC, van Overeem Hansen T. (2011) The BTNL2 A allele variant is frequent in danish patients with sarcoidosis. Clin Respir J 5: 105-111. 10.1111/j.1752-699X.2010.00206.x; 10.1111/j.1752-699X.2010.00206.x.
307. Coudurier M, Freymond N, Aissaoui S, Calender A, Pacheco Y, et al. (2009) Homozygous variant rs2076530 of BTNL2 and familial sarcoidosis. Sarcoidosis Vasc Diffuse Lung Dis 26: 162-166.
308. Arnett HA, Escobar SS, Gonzalez-Suarez E, Budelsky AL, Steffen LA, et al. (2007) BTNL2, a butyrophilin/B7-like molecule, is a negative costimulatory molecule modulated in intestinal inflammation. J Immunol 178: 1523-1533.
309. Cozier Y, Ruiz-Narvaez E, McKinnon C, Berman J, Rosenberg L, et al. (2013) Replication of genetic loci for sarcoidosis in US black women: Data from the black women's health study. Hum Genet 132: 803-810. 10.1007/s00439-013-1292-5; 10.1007/s00439-013-1292-5.
310. Mitsunaga S, Hosomichi K, Okudaira Y, Nakaoka H, Kunii N, et al. (2013) Exome sequencing identifies novel rheumatoid arthritis-susceptible variants in the BTNL2. J Hum Genet . 10.1038/jhg.2013.2; 10.1038/jhg.2013.2.
311. Brand O, Gough S, Heward J. (2005) HLA , CTLA-4 and PTPN22 : The shared genetic master-key to autoimmunity? Expert Rev Mol Med 7: 1-15. 10.1017/S1462399405009981.
312. Goris A, Liston A. (2012) The immunogenetic architecture of autoimmune disease. Cold Spring Harb Perspect Biol 4: 10.1101/cshperspect.a007260. 10.1101/cshperspect.a007260; 10.1101/cshperspect.a007260.
313. Andreatta M, Nielsen M. (2012) Characterizing the binding motifs of 11 common human HLA-DP and HLA-DQ molecules using NNAlign. Immunology 136: 306-311. 10.1111/j.1365-2567.2012.03579.x; 10.1111/j.1365-2567.2012.03579.x.
314. Sidney J, Steen A, Moore C, Ngo S, Chung J, et al. (2010) Five HLA-DP molecules frequently expressed in the worldwide human population share a common HLA supertypic binding specificity. J Immunol 184: 2492-2503. 10.4049/jimmunol.0903655.
315. Wahlstrom J, Dengjel J, Winqvist O, Targoff I, Persson B, et al. (2009) Autoimmune T cell responses to antigenic peptides presented by bronchoalveolar lavage cell HLA-DR molecules in sarcoidosis. Clin Immunol 133: 353-363. 10.1016/j.clim.2009.08.008.
316. Gregersen PK, Silver J, Winchester RJ. (1987) The shared epitope hypothesis. an approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum 30: 1205-1213.
317. Castelli FA, Buhot C, Sanson A, Zarour H, Pouvelle-Moratille S, et al. (2002) HLA-DP4, the most frequent HLA II molecule, defines a new supertype of peptide-binding specificity. J Immunol 169: 6928-6934.
318. Achkar JP, Klei L, de Bakker PI, Bellone G, Rebert N, et al. (2012) Amino acid position 11 of HLA-DRbeta1 is a major determinant of chromosome 6p association with ulcerative colitis. Genes Immun 13: 245-252. 10.1038/gene.2011.79; 10.1038/gene.2011.79.
319. Valente FP, Tan CR, Temple SE, Phipps M, Witt CS, et al. (2009) The evolution and diversity of TNF block haplotypes in european, asian and australian aboriginal populations. Genes Immun 10: 607-615. 10.1038/gene.2009.45; 10.1038/gene.2009.45.
320. Hollenbach JA, Holcomb C, Hurley CK, Mabdouly A, Maiers M, et al. (2013) 16(th) IHIW: Immunogenomic data-management methods. report from the immunogenomic data analysis working group (IDAWG). Int J Immunogenet 40: 46-53. 10.1111/iji.12026; 10.1111/iji.12026.
321. Gourraud PA, Meenagh A, Cambon-Thomsen A, Middleton D. (2010) Linkage disequilibrium organization of the human KIR superlocus: Implications for KIR data analyses. Immunogenetics 62: 729-740. 10.1007/s00251-010-0478-4; 10.1007/s00251-010-0478-4.

322. Yang X, Chockalingam SP, Aluru S. (2013) A survey of error-correction methods for next-generation sequencing. Brief Bioinform 14: 56-66. 10.1093/bib/bbs015; 10.1093/bib/bbs015.
323. Dilthey A, Leslie S, Moutsianas L, Shen J, Cox C, et al. (2013) Multi-population classical HLA type imputation. PLoS Comput Biol 9: e1002877. 10.1371/journal.pcbi.1002877; 10.1371/journal.pcbi.1002877.
324. Gorer PA, Schutze H. (1938) Genetical studies on immunity in mice: II. correlation between antibody formation and resistance. J Hyg (Lond) 38: 647-662.
325. Dausset J. (1958) Iso-leuko-antibodies. Acta Haematol 20: 156-166.
326. Lee HM, Hume DM, Vredevoe DL, Mickey MR, Terasaki PI. (1967) Serotyping for homotransplantation. IX. evaluation of leukocyte antigen matching with the clinical course and rejection types. Transplantation 5: Suppl:1040-5.
327. Mack SJ, Sanchez-Mazas A, Single RM, Meyer D, Hill J, et al. (2007) Population samples and genotyping technology. Tissue Antigens 69 Suppl 1: 188-191. 10.1111/j.1399-0039.2006.00768.