

**Identification of sarcoidosis susceptibility genes by association
mapping and candidate gene approaches**

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität
zu Kiel

vorgelegt von

Ruta Valentonyte, M.sc

Kiel

2004

Referent/in:.....

Koreferent/in.....

Tag der mündlichen Prüfung:.....

Zum Druck genehmigt: Kiel,

Der Dekan

Table of contents

Table of contents	3
Abbreviations and symbols	6
Amino acid symbols.....	9
DNA bases.....	9
1. Introduction	10
1.1. Susceptibility gene finding in complex diseases	10
1.1.1. Linkage analyses	10
1.1.2. Linkage disequilibrium in the MHC and entire human genome	11
1.1.3. Association studies	12
1.1.4. Candidate gene analyses	13
1.2. Sarcoidosis.....	14
1.2.1. The disease	14
1.2.2. Familial aggregation of sarcoidosis	15
1.2.3. Sarcoidosis and other hereditary granulomatous disorders.....	16
1.2.4. Sarcoidosis linkage studies	16
1.2.5. Human Major Histocompatibility Complex.....	18
1.2.6. Sarcoidosis associations with HLA genes.....	19
1.2.7. Autoimmune and chronic inflammatory disease mapping to the human MHC.....	20
1.2.8. Sarcoidosis candidate genes in the entire genome	23
1.3. Aims of the study.....	25
2. Materials and methods	26
2.1. Materials and equipment	26
2.2. Electronic data bases	29
2.3. Participants	30
2.3.1. Sarcoidosis population for the chromosome 6 association study.....	30
2.3.2. Control group individuals	31
2.3.3. Sequencing patients for association mapping experiment on chromosome 6..	32
2.3.4. CARD15 gene analysis samples	32
2.4. DNA samples	33
2.4.1. Isolation of genomic DNA from blood	33
2.4.2. Plates for genotyping with TaqMan assays	33
2.4.3. HLA-DRB1 typing plate design	34

2.5.	Di-allelic genotyping.....	35
2.5.1.	Di-allelic assay design and genotyping.....	37
2.5.2.	SNP selection for chromosome 6 mapping	40
2.5.3.	Typing of the <i>CARD15</i> gene.....	40
2.6.	Multi-allelic genotyping	41
2.6.1.	The Problem of multi-allelic genotyping	41
2.6.2.	Genotyping of HLA-DRB1 by allele -group PCR.....	41
2.6.3.	Development of a non-electrophoretic method for high-throughput DRB1 allele-group identification	44
2.7.	SNP verification and mutation detection	48
2.7.1.	SNP selection.....	48
2.7.2.	Primer design and PCR	49
2.7.3.	Sequencing	50
2.7.4.	SNP verification and mutation detection in the association region and in sarcoidosis candidate genes	52
2.7.4.1.	The <i>BTNL2</i> locus on chromosome 6	52
2.7.4.2.	<i>CARD15</i> on chromosome 16	52
2.8.	Internal data bases	53
2.9.	Statistical analysis	54
3.	Results	56
3.1.	Association study in the peri-MHC	56
3.1.1.	Identification of the association lead on chromosome 6.....	56
3.1.2.	Analysis of LD in the screened region	60
3.1.3.	High density SNP mapping in the association region	62
3.1.4.	Separation of the genetic effect of <i>BTNL2</i> and the HLA-DRB1 genes	68
3.1.5.	<i>BTNL2</i> structural analysis.....	71
3.1.6.	Functional exploration of <i>BTNL2</i> mutations	72
3.2.	Results of the <i>CARD15</i> gene analysis	74
3.2.1.	<i>CARD15</i> mutations in sarcoidosis	74
3.2.2.	Sequence analysis of the <i>CARD15</i> gene NOD domain	77
4.	Discussion.....	78
4.1.	Association of the <i>BTNL2</i> mutations with sarcoidosis	78
4.1.1.	Experimental design.....	78
4.1.2.	Identification of the association region on chromosome 6p.....	78

4.1.3.	Sarcoidosis candidate genes on chromosome 6p.....	79
4.1.4.	Genotyping of the HLA-DRB1	82
4.2.	CARD15 mutations in sarcoidosis	85
5.	Conclusions	89
6.	Summary: Identification of sarcoidosis susceptibility genes by association mapping and candidate gene approaches	91
7.	Zusammenfassung: Identifikation von Sarkoidose Krankheitsgenen durch Assoziationsstudien und Kandidatengen Ansätze.....	93
8.	References	95
9.	Appendix	108
9.1.	Supplemental data	108
9.1.1.	Primer and probe information for chromosome 6 assays	108
9.1.2.	Sarcoidosis patient phenotypes.....	112
9.1.3.	Structure of 138 sarcoidosis families investigated in the CARD15 study	113
9.1.4.	Assays for CARD15 genotyping	114
9.1.5.	Reaction conditions for diallelic genotyping assays	115
9.1.6.	Mutation detection primers used for chromosome 6 PCR and sequencing experiments.....	120
9.1.7.	Sequence variations detected via sequencing	122
9.1.8.	Sequence data	126
9.2.	Additional experiments on BTNL2	128
9.2.1.	Preparation of Bronchoalveolar lavage cell samples	128
9.2.2.	BTNL2 expression experiment.....	129
9.2.3.	Verification of the splicing effects of rs2076530.....	130
9.2.4.	In silico BTNL2 protein analysis	131
9.2.5.	Computational 3D structure model of the second IgV domain and the following IgC domain of the BTNL2 homodimer	132
	Index of figures.....	133
	Index of tables	134
10.	Curriculum vitae	135
11.	Declaration (Erklärung) and publication list	137
12.	Acknowledgements	139

Abbreviations and symbols

AbD	assay(s) by design (Applied Biosystems, available at https://www.store.appliedbiosystems.com)
AGER	advanced glycosylation end-product-specific receptor (synonym RAGE)
AoD	assay(s) on demand (Applied Biosystems, available at https://www.store.appliedbiosystems.com)
ASP	affected sibling pair
B7-1	B7-1 antigen
BAC	bacterial artificial chromosome
bp.	base pair(s)
BTNL2	butyrophilin-like 2 (alternative symbol BTL-II)
°C	degree Celsius
CARD15	caspase recruitment domain family, member 15 (alternative symbol NOD2)
CCR2	chemokine (C-C motif) receptor 2
CD4+	T cells that recognise their specific antigens in association with major histocompatibility complex class II molecules
CD8+	T cells that recognise their specific antigens in association with major histocompatibility complex class II molecules
CD28	CD28 antigen
CD80	CD80 antigen (B7-1 antigen)
CEPH	Centre d'Etude du Polymorphisme Humain (cell lines)
CTLA4	cytotoxic T-lymphocyte-associated protein 4
DDW	double distilled water
DNA(s)	deoxyribonucleic acid(s)
dNTP	2' -deoxynucleoside -5' -triphosphate
DQA1	see HLA-DQA1
DRA	see HLA-DRA
DRB1	see HLA-DRB1
dsDNA	double stranded DNA
e.g.	exempli gratia
EM	expectation maximisation
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1
HLA-DRA	major histocompatibility complex, class II, DR alpha

HLA-DRB1	major histocompatibility complex, class II, DR beta 1
FAM TM	6-carboxyfluorescein
Fig(s).	figure(s)
HLA	human leukocyte antigen
HWE	Hardy-Weinberg-Equilibrium
i.e.	id est
IgC	immunoglobulin constant domain
IgV	immunoglobulin variable domain
kb	kilo base(s)
LD	linkage disequilibrium
LTA	lymphotoxin alpha
MB	mega base(s)
MHC	human major histocompatibility complex
MICA	MHC class I polypeptide-related sequence A
MICB	MHC class I polypeptide-related sequence B
μl	microlitre(s)
mmol	millimole
NCBI	National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov)
nmol	nanomole(s)
NOD	nucleotide oligomerisation domain
NOTCH4	Notch homolog 4 (Drosophila)
NPL	nonparametric linkage
PAR	population attributable risk
pmol	picomole(s)
Rn	fluorescent emission of the normalised reporter dye
RR	familial relative risk
TAMRA	6-carboxytetramethylrhodamine
TAP2	transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)
TaqMan®	commercial name for sequence variation detection assay, using 5' 3' exonuclease activity
TaqMan®-MGB-	commercial name for sequence variation detection assay, using 5' 3' exonuclease activity and 3' minor groove binding probes
TCR	T cell receptor
TDT	transmission disequilibrium test

TE	Tris-hydroxymethyl aminomethane buffer
TET	tetachloro-6-carboxyfluorescein
THP-1	myelomonocytic cell line
Tm	melting temperature
TNF α	tumor necrosis factor alpha (TNF superfamily, member 2)
TSBP	testis specific basic protein
SD	standard deviation
SNP(s)	single nucleotide polymorphism(s)
SPSS	statistical package for the social sciences
SSPCR	sequence specific polymerase cyclic reaction
STR	short tandem repeat(s)
VIC®	Applied Biosystems trade name for fluorescent dye
Yrs	years

Amino acid symbols

A	Alanine
C	Cysteine
D	Aspartic acid
E	Glutamic acid
F	Phenylalanine
G	Glycine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Threonine
V	Valine
W	Tryptophan
Y	Tyrosine

DNA bases

A	Adenine
G	Guanine
C	Cytosine
T	Thymine

1. Introduction

1.1. Susceptibility gene finding in complex diseases

Linkage and association studies represent the two major approaches that attempt to determine the genetic loci that confer susceptibility to diseases with a complex genetic background. Linkage studies attempt to identify the approximate genetic location of a disease trait. Association studies provide a method for the fine localisation. Association tests can be pedigree or singleton-based. A case-control study is population based and investigates the association between a disease trait and genetic markers using data (different genotype frequencies) from unrelated patients (singletons) and unrelated controls. Genetic linkage analysis tests for co-segregation between a marker (the most common used markers are short tandem repeats (STRs) which are often referred to as microsatellites) and a putative disease locus. Genotype association is useful to detect the precise location of disease susceptibility markers. Such positional candidates are then investigated to characterise their potential function in the context of the disease.

1.1.1. Linkage analyses

In complex diseases, linkage analysis tests are performed to detect a co-segregation between a marker and a putative disease locus. The traditional lod score approach ¹ tests the likelihood of reduced recombination between a marker locus and disease gene versus independent segregation of the two loci. Generally, highly-polymorphic STRs, whose chromosomal locations are known, are genotyped using the polymerase chain reaction with the resulting fragments separated electrophoretically according to length. Such microsatellites are distributed throughout the genome at around 100kb intervals and the different alleles vary in repeat length ².

Parametric lod score analysis requires a considerable number of large families and a clear Mendelian mode of inheritance to be valid to show genetic linkage. However, in complex diseases, pedigrees are generally small and the mode of inheritance tends to be unclear. This limitation has led to an emphasis on non-parametric linkage (NPL) analysis methods, which do not require the mode of inheritance to be specified. Such analyses determine if allele transmission in affected siblings pairs (ASP) deviate from Mendelian expectations ³. A

significantly increased level of allele sharing (alleles that are identical by descent) suggests the marker is linked to the disease locus ⁴. Multipoint linkage analysis is performed testing at the same time several markers located on the same chromosome. The closer the microsatellite is located to the disease gene, the higher LOD score value is expected because of rarer recombination. Successful location of linkage depends much on informativity of markers, sample size and phenotype complexity ⁵

1.1.2. Linkage disequilibrium in the MHC and entire human genome

Each new allele that arises by mutation in the sequence is initially associated with the particular chromosomal background on which it arose. A set of closely linked alleles (genes or DNA polymorphisms) inherited as a unit is called haplotype. Different combinations of polymorphisms are known as haplotypes. Coinheritance of haplotypes leads to associations between the alleles present on the haplotypes at the population level ⁶. This phenomenon of non-random association between alleles at different but adjacent loci is known as linkage disequilibrium (LD) ⁷. A common measure of LD is the absolute D' value ⁸, which is a standardization of the LD coefficient D to an index between zero and one corresponding to no LD and complete LD respectively. D is the difference between the observed and the expected haplotype frequencies at two diallelic loci ⁹.

Investigations of LD suggest that the genome consists of discrete haplotype blocks in which there is a little evidence of historical recombination ¹⁰⁻¹². LD blocks may extend from a few kilobases (kb) to regions greater than 100 kb ¹³. A useful measure is the „half-length” of LD (the distance at which the average $|D'|$ drops below 0.5). For example, LD in a United States population of northern-European has a half-length of about ~60 kb and in a Nigerian population extends markedly less far ¹³. New haplotypes emerge through novel mutation events and through recombination between homologous sites in the paternal and maternal chromosomes ⁶. LD patterns are influenced not only by „hot” and „warm” spots of recombination but also by selective pressures and demographic events ^{7, 14, 15}.

Haplotype-based methods offer a powerful approach to disease gene mapping, based on association between causal mutations and the ancestral haplotype ¹¹. Newly introduced mutations are in LD with the alleles on the ancestral haplotype. Thus, genotyping effort and costs might be effectively reduced by genotyping only a small fraction of carefully selected

SNPs that tag common haplotypes^{10, 16}. Further steps would then include a fine mapping experiment in the region of strong LD¹². On the other hand, strong association between nearby alleles can make identification of the causal variant very difficult. The statistical power to detect an allelic association depends on specific properties of each marker, including its frequency and population history¹⁰.

The organization of the Human Major Histocompatibility Complex (MHC), which has been implicated in many immune disorders, appears to be well conserved. The regions encoding the class I and II molecules in most of the cases (~97%) are inherited without recombination¹⁷. Investigations into MHC evolution suggest that the region is derived from a single ancestral gene, which has expanded through serial duplications and has been modified by mutations and contracted by deletions. Selection pressures have accompanied all these processes^{18,530}. Although classical human-leukocyte-antigen (*HLA*) genes in this region represent good candidates for immune disorders, linkage disequilibrium makes it difficult to rule out the influence of neighbouring genes on disease susceptibility. The few studies that have analysed the haplotype structure of the MHC region have suggested that LD variation in the MHC, with the exception of the classical *HLA* loci¹⁹, is not essentially different from that in the rest of the genome^{14, 20-22}.

1.1.3. Association studies

The main use of association tests in genetics is to identify genes that possibly influence disease susceptibility. Disease association tests are carried out using either a case-control or family-based study design. Single nucleotide polymorphisms (SNPs) are used to localise disease susceptibility genes.

Case-control studies use data from unrelated affected individuals and unrelated controls. Evidence for a statistically significant association between a marker and the disease is taken from a comparison of the genotype (and thus allele) frequencies in both samples. A major limitation with case-control association studies is that an association may arise if cases and controls were drawn from genetically different populations even though no disease gene association exists. Admixture in populations lead to a potential excess of both false positive and false negative associations. The possibility of false associations depends on the population sampled, the trait being studied and the marker being tested^{23, 24}.

Family-based association studies use data from within families. Limiting analysis to familial genetic information helps to avoid the matching problems experienced in case-control studies. Sampling families is generally more difficult than cases and controls, but this methodology is still possible in a sarcoidosis study because sarcoidosis typically has an onset in early-adulthood and thus both parents of the proband may be available²⁴. The association test in pedigrees is performed measuring distorted transmission of alleles²⁵.

Sarcoidosis literature is replete with case-control association studies. However, the findings often conflict with one another. One potential explanation for inconsistent results is that the same phenotype can be caused by damage of different genetic mechanisms²⁶. Discrepancies among studies can also arise due to differing sampling methodologies²⁴. Ideally, in the functional candidate gene approach, associations found through case-control studies should be confirmed in family-based studies²⁷.

1.1.4. Candidate gene analyses

The selection of candidate genes includes a large number of potential factors that are most likely to be involved in the phenotype. Knowledge about biology, disease pathophysiology, biological plausibility are evaluated. Functional effects of single nucleotide polymorphisms (SNPs) on the protein are considered. Multiple SNPs can contribute to association, thus the ones with strongest association are identified. Dose effect is evaluated in cases, when numbers of variant copies contribute to a greater risk^{24, 28}.

Information about the location and type of the sequence variants in a gene can be used to prioritize polymorphisms²⁸. Nonsense, missense, splice site mutations, and deletions in the coding sequence can affect protein structure. Synonymous changes do not primarily affect the protein structure. However, they may influence the mRNA splicing, if the polymorphism is located in an exonic splicing enhancer^{29, 30}. Mutations affecting gene expression levels are usually localised in promoter, intronic and intergenic regions. Interpretation of the latter mutations is more complex. As polymorphisms are frequent in the human genome (approximately 1 in 1,000 bp), it is important to select carefully a limited number of SNPs to genotype from the set that is often available in a particular candidate gene^{28, 31}.

Replication of association is another critical aspect. Inconsistency of results between different studies and different populations might point to the need for more studies in certain populations or more detailed study of the function of a candidate gene ²⁸. However, population diversity may also be a reason of lack of replication ³² (e.g. CARD15 is associated with Crohn's disease in European population, but not in Asians) ²⁶

1.2. Sarcoidosis

1.2.1. The disease

Sarcoidosis is a systemic inflammatory disease of unknown aetiology, characterized by T-lymphocyte and mononuclear phagocyte infiltration leading to granuloma formation ^{33,34}. The hallmark of the disease is the presence of noncaseating granulomas in affected organs. These granulomas can occur with varying rates in any organ system, but most commonly are found in the lung and lymph nodes ³³. A range of 30-60% of reported cases of sarcoidosis are asymptomatic and have been discovered by the presence of characteristic findings on routine health screening chest radiographs ³⁵⁻³⁷. For symptomatic patients, the most common problems include respiratory symptoms e.g. dry cough, dyspnea, chest pain and systemic symptoms e.g. fever, night-sweats, fatigue and malaise. The characteristic granuloma found in sarcoidosis is discrete and compact, composed of multinucleated giant cells, mononuclear phagocytes and lymphocytes surrounded by tightly organized fibroblasts, mast cells and an extracellular matrix ³⁸. The finding of noncaseating granulomas in biopsy specimens requires the presence of correlating clinical symptoms consistent with sarcoidosis and the exclusion of alternative diseases like tuberculosis to establish confidently the diagnosis.

Clinical presentation of the disorder varies widely, with two main phenotypes: acute and chronic sarcoidosis. Acute sarcoidosis resolves within two to three years, often disappearing spontaneously and leaving no residual effects. A subset of acute patients exhibit Löfgren's syndrome, which is characterised by a combination of fever, bilateral hilar lymphadenopathy, erythema nodosum and arthritis, often at the ankle joints. Approximately one-quarter of sarcoidosis patients have permanent lung damage because of scarring, and some may suffer from chronic sarcoidosis. In about five percent of patients, the damage from sarcoidosis progresses enough that organ transplantation (such as lungs, heart, liver) is required or death occurs ³⁹⁻⁴².

Despite of evidence suggesting that sarcoidosis results from a complex interaction of multiple genes with environmental exposures or infection, no specific aetiologic agent has yet been identified. Disease presentation in certain seasons supports the idea of an infective trigger, possibly microbial in origin. The initiating agent responsible for sarcoidosis is likely to be airborne since the lungs are the most commonly affected organ in this complex disease^{43, 44}. The seasonal peak occurs during spring⁴⁵. Investigations of bacteria (e.g. mycobacteria), viruses (e.g. herpes, Epstein-Barr, retrovirus, cytomegalovirus) and environmental organic (e.g. pine tree pollen, clay) or inorganic (e.g. aluminium, zirconium, talc) antigens have not consistently demonstrated an association with sarcoidosis⁴⁶⁻⁵⁰.

Many individuals with sarcoidosis are characterised as asymptomatic. Therefore, estimates of the reported prevalence may depend on the way in which the epidemiological data is generated. The reported prevalence varies significantly by sex, racial group, and country. Nevertheless, sarcoidosis appears to be one of the most frequent interstitial lung diseases⁵¹. There are up to 64 patients per 100,000 individuals in Sweden to 9 per 100,000 in Italy. Intermediate numbers are observed in Denmark (53 per 100,000), Germany (43), Ireland (40), Norway (27), UK (20) and France (9). Within the United States, the prevalence for the Caucasian population of North America is 3 and for Afro-Americans 47 per 100,000^{34, 52}. Sarcoidosis is found in all races, affecting slightly more females than males. It most commonly affects individuals aged 20 to 40 years. Initial presentation of sarcoidosis is also related to race- African-Americans are at greater risk of morbidity and mortality^{36, 53}.

1.2.2. Familial aggregation of sarcoidosis

Numerous reports on the familial occurrence of sarcoidosis suggest a genetic component to disease predisposition^{54, 55}. Data from a nationwide chest radiography screening program indicated a recurrent risk in siblings of affected patients of approximately one in 100, which is 20 times higher than the prevalence rate of sarcoidosis in Germany⁵⁶. A familial relative risk observed in American whites (RR=18.0) was markedly higher compared with African-American (RR=2.8) cases⁵⁴. Monozygotic twins are 24 times more concordant for disease than dizygotic twins⁵⁷. From these studies, it is clear, that family members have a several-fold increased risk of disease compared with the general population. The familial sarcoidosis studies clearly exclude a simple mode of inheritance, suggesting a more complex genetic

background. Genetic mechanisms such as single gene effects with greatly reduced penetrance, polygenic inheritance or a multifactorial background with interactions between genetic make-up and environmental agents could be characteristic features of sarcoidosis⁵⁸.

1.2.3. Sarcoidosis and other hereditary granulomatous disorders

Sarcoidosis demonstrates remarkable overlap with other chronic inflammatory disorders. Blau syndrome and Crohn's disease are complex disorders that show overlap with sarcoidosis in the spectrum of involved organs and their immunopathophysiology.

Enhanced immune response to a yet unknown target with monocyte and T-cell activation and granuloma formation in affected tissues are characteristic features of both sarcoidosis and Crohn's disorders^{33, 34, 59}. Although the primary sites of inflammation differ (lung in sarcoidosis and intestine in Crohn's disease), affection of skin, joints and other organs is common in both disorders. Coincidence of sarcoidosis and Crohn's disease in families or even in one individual is more common than expected by chance alone⁴².

Blau syndrome demonstrates clinical similarities with sarcoidosis. The disease is characterised by granulomatous arthritis, uveitis and erythema⁶⁰. Blau syndrome is caused by a single dominant gene mutation and has onset in early childhood. Based on clinical similarities with sarcoidosis, Blau syndrome might be a monogenic early-onset variant of sarcoidosis.

1.2.4. Sarcoidosis linkage studies

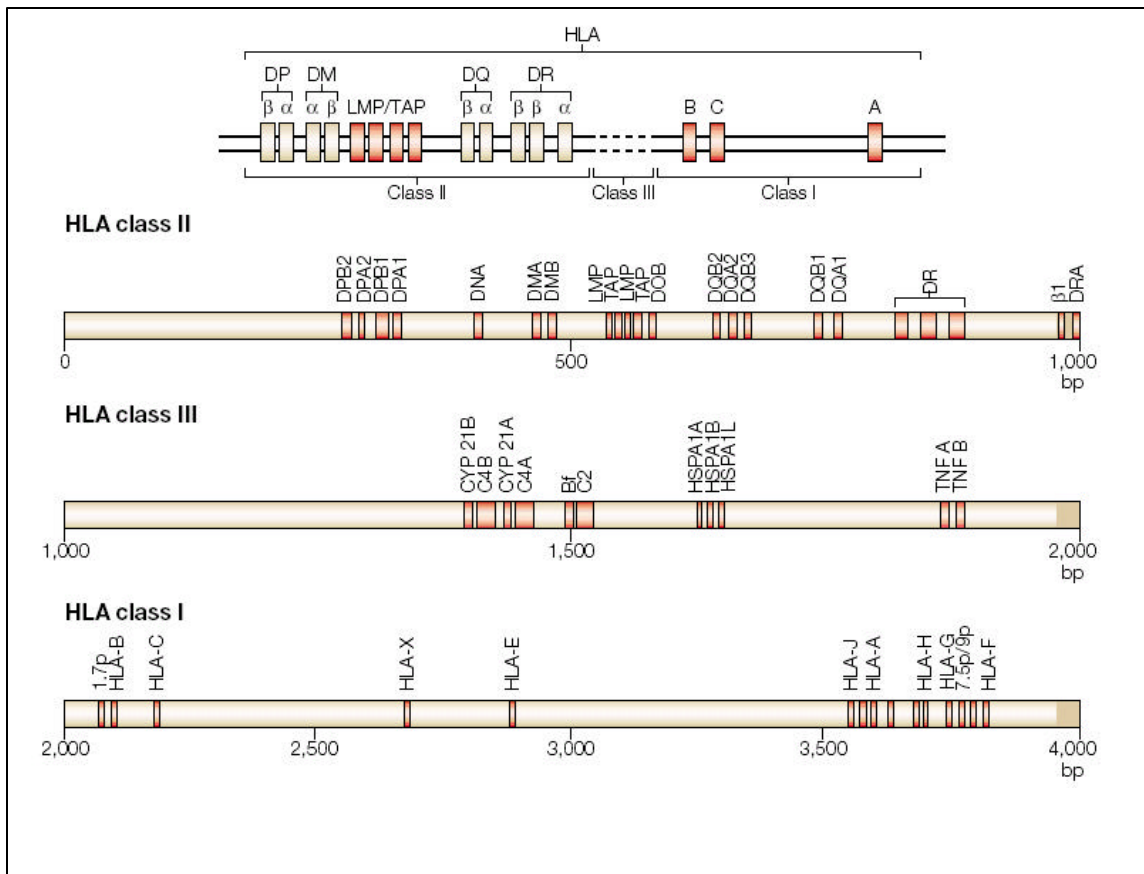
In order to localize sarcoidosis susceptibility genes, several linkage studies and many association studies have been performed in different populations. A linkage study of seven polymorphisms spanning the MHC region in 55 German sarcoidosis families gave an estimated maximum NPL score of >2.5 for the entire MHC region, with a maximum value of 3.2 at marker locus D6S1666 (panel A of fig.1)⁶¹(see NPL definition in chapter 1.1.1). Later, a detailed genome-wide linkage study of 63 German sarcoidosis families with 225 microsatellite markers identified seven peaks of linkage evidence located on six chromosomes. The most prominent peak was still located on chromosome 6 at the

located 3.9Mb centromeric from the marker D6S1666 which was implicated as indicating a susceptibility region in German families.

1.2.5. Human Major Histocompatibility Complex

The human Major Histocompatibility Complex (MHC) is located on chromosome 6p21 and is composed of three classes: (see fig.2) MHC class I, II and III. MHC class I mainly includes the *HLA-A*, *HLA-B* and *HLA-C* genes, MHC class II contains the *HLA-DR*, *HLA-DQ* and *HLA-DP* subclasses. The primary immune relevant genes in the MHC class III region (located between the class I and class II regions) encode TNF-alpha, TNF-beta, the complement proteins (C4A, C4B, C2 and Bf), enzymes involved in steroid synthesis (CYP21A and CYP21B) and heat-shock proteins (HSPA1A; HSPA1B and HSPA1L) ⁶³.

Fig.2. The Major histocompatibility complex (published by ⁶³). Upper panel: classification of the region into three classes. Lower panel: HLA gene content in every subclass.



The products of the Human Leucocyte Antigen (*HLA*) genes play a fundamental role in the regulation of immune processes by antigen presentation to T cells. The *HLA* genes and their products are grouped into two classes according to their chemical structure and biological properties. Class I molecules contain one heavy chain (45 kD) and a light chain called β 2-microglobulin (12 kD) that contributes to the overall structure of the protein. Class II molecules do not contain β 2-microglobulin and consist of two (alpha and β) chains of similar size (34 and 30 kD). Both classes differ according to their peptide-binding groove, which accommodates 8-9 amino acids in class I but is open-ended in class II. MHC class I molecules present endogenously amplified antigens, e.g. viral proteins, to CD8+ T-cells. MHC class II molecules present exogenously derived proteins, e.g. bacterial products or viral capsid proteins, to CD4+ T cells. However, if the immune mechanism malfunctions and T-cells react against “self”-peptides, an autoimmune disease may arise. MHC class I molecules are widely expressed, although the level varies between different cell types. MHC class II molecules are constitutively expressed only by certain cells involved in immune responses, though they can be induced in a wider variety of cells. All *HLA* alleles are expressed co-dominantly. In about 97% of individuals, the entire MHC haplotype is inherited without recombination. Each *HLA* gene has an unusual number of alleles (for instance *HLA-DRB1* gene has over 200 different alleles) caused by high levels of polymorphism in the region.^{176, 177}

1.2.6. Sarcoidosis associations with HLA genes

Many *HLA* genes have been tested for association with sarcoidosis; investigating the theory that disease-associated *HLA* molecules present specific antigenic peptides in such a way that the recognizing T-lymphocytes initiate an inflammatory response with peculiar pathological consequences⁶⁴. Originally, associations with sarcoidosis were attributed to the MHC-I alleles. A positive association between *HLA-B7* and *HLA-B13* was found in Moravian Czechs⁶⁵. However, the *HLA-B7* was significantly decreased in Japanese patients⁶⁶.

Recently, many studies have concentrated on MHC-II typing sustaining the hypothesis that sarcoidosis is triggered by exogenous antigens. *HLA-DR* allele typing have provided comprehensive results for numerous populations, although reports of risk alleles were conflicting. *HLA-DR5*, *HLA-DR8* and *HLA-DR9* were risk alleles in Japanese⁶⁶⁻⁶⁸. *HLA-DR5* has been found significantly associated with chronic form of sarcoidosis in Germans⁶⁹.⁷⁰ *HLA-DR3* was associated with acute disease in Germans and Scandinavians, whilst the

chronic form in Scandinavians was associated with HLA-DR14 and HLA-DR15⁶⁹⁻⁷². A study in Asian Indians reported two risk alleles: HLA-DRB*11 and HLA-DRB1*14⁷³. Recently, significant association with HLA-DRB1*1101 alleles was found in American blacks and whites⁷⁴. More consistent results were achieved when evaluating „protective” *HLA-DRB* alleles in different populations. The HLA-DR alleles, which encode DR1 and DR4 antigens, which share characteristic small hydrophobic residues at amino acid position 11, were negatively associated with Scandinavian, Japanese, British, Polish and Czech sarcoidosis cases⁷². In this study, a previously reported association between sarcoidosis and single nucleotide polymorphisms of the *TAP2* gene⁷⁵ was shown to be probably due to linkage disequilibrium with the *HLA-DR* locus.

HLA-DQ locus analysis has provided fewer significant alleles. HLA-DQA1 0101/4 was significantly associated with sarcoidosis in Asian Indians⁷³. Alleles of the more centromeric gene HLA-DQB1*0201 and HLA-DQB1*0602 were over-transmitted in African-American sarcoidosis families⁷⁶. However, the HLA-DQB1*0201 was strongly „protective” against severe sarcoidosis in British, Dutch and Asian Indian patients, whereas HLA-DQB1*0602 was associated with disease^{73, 77}. HLA-DQB1*0503 has been found associated with disease in Asian Indians.

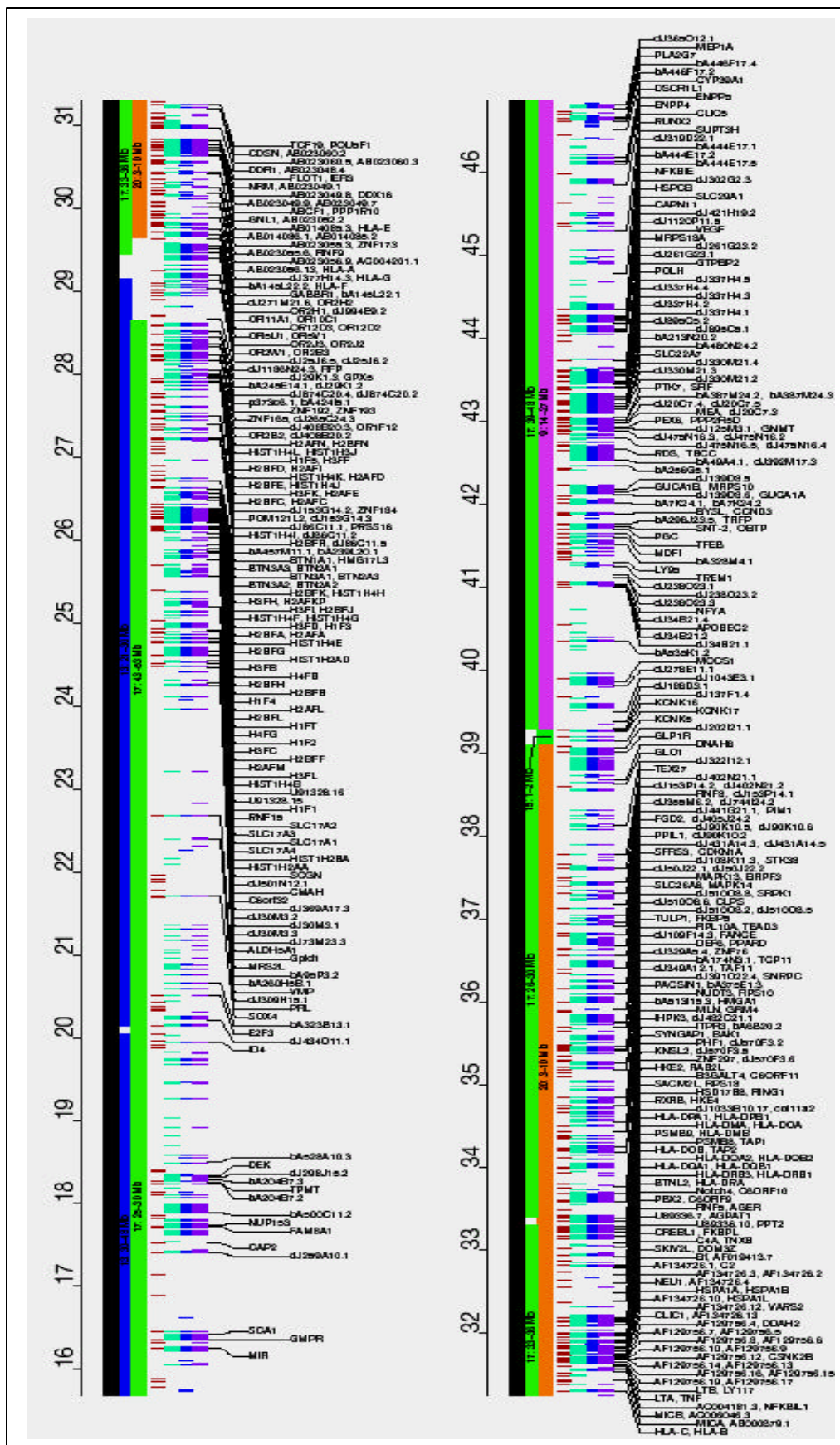
Despite only few reports about the *HLA-DP* locus, the *HLA-DPB1* gene is interesting because of its association with berylliosis (a phenocopy of sarcoidosis characterised by accumulation of beryllium-specific CD4+ T lymphocytes). An allele-association study indicated that a glutamine located at position 69 of DPB1 confers susceptibility to chronic beryllium disease⁷⁸. The immunopathologic and clinical similarities between chronic beryllium disease and sarcoidosis suggest that similar immune-response genes may be involved in susceptibility to both diseases. However, no association was found with the Glu69+ allele in sarcoidosis patients from the US⁷⁹ and from Europe^{61, 75}. Further HLA-DPB1 amino acids, Val36+ and Asp55+, which are involved in antigen binding have been shown to be significantly associated with sarcoidosis risk in African-American cases⁷⁹.

1.2.7. Autoimmune and chronic inflammatory disease mapping to the human MHC

Numerous studies have indicated a role for the peri-MHC in susceptibility to chronic inflammatory and autoimmune diseases. Most studies have focused on the allelic variants of

the *HLA* genes, as the products of these genes are primarily involved in immune processes. Autoimmune responses involve T-cells, which respond to a particular MHC genotype-dependent antigen. Thus, the associations are supported by the hypothesis that susceptibility to autoimmune disease could be caused by certain *HLA* allelic variants through their presentation of auto-antigenic peptides to autoreactive T-cells. However, there are many other inflammatory response related genes in the MHC region (for gene content in the peri-MHC region see fig.3), which have not been tested for disease association. Therefore, the MHC region is still one of the main targets for association studies for autoimmune disorders.

Fig.3. Gene content in peri-MHC region (published by ⁸⁰)



Linkage and association studies on chromosome 6p have been performed on patients with inflammatory bowel disease ⁸¹, psoriasis ⁸², systemic lupus erythematosus ⁸³, insuline-dependent diabetes ⁸⁴, rheumatoid arthritis ⁸⁵ and multiple sclerosis ⁸⁶. All these studies indicated the presence of genetic factors located in the MHC.

Many association studies have investigated non-HLA genes as well that map to the MHC. However, association results obtained from investigations of Caucasian sub-populations or different populations often do not overlap or are even contradictory. Thus, in only a few successful studies an initial association between a typical *HLA* gene and a disease been excluded by finding a nearby locus possessing a causative mutation. Mutations in the *CYP21A2* gene were found in patients suffering from congenital adrenal hyperplasia ^{87, 88}. Defects of the *HFE* gene were assigned to the iron storage disorder referred as hereditary haemochromatosis ⁸⁹. Sequence variants in the lymphotoxin-alpha (*LTA*) gene were indicated as risk factors for myocardial infarction ⁹⁰.

1.2.8. Sarcoidosis candidate genes in the entire genome

Numerous reports have implicated angiotensin-converting enzyme (ACE) in the pathophysiology of sarcoidosis. The gene has a 300bp insertion/deletion (I/D) polymorphism, which influences serum ACE levels. ID and DD genotypes were found to be associated with the disease in several studies ⁹¹⁻⁹³. However, in later studies no co-segregation was found with familial sarcoidosis ⁹⁴ and no associations with the various sub-phenotypes of the disease was detected in a case-control study in two European populations ⁹⁵. Therefore, it seems that *ACE* polymorphisms do not represent a major hereditary cause of the disease, but rather, they modify the course of the disorder ⁹⁴.

The inflammatory response in sarcoidosis is characterised by an increased production of several pro-inflammatory cytokines and chemokines at the sites of disease ³⁴. Therefore, it is plausible that polymorphisms in these genes may contribute to differences in the regulation of cytokine release ⁹⁶. In sarcoidosis major mediators are released by both T-lymphocytes (IL-2, IFN-gamma, LT) and alveolar macrophages (IL-1beta, IL-12, TNF-alpha). Tumor necrosis factor –alpha (TNFa) is thought to play a key role as a mediator of inflammation and cellular response. TNFa production is increased in the sarcoid lung and its role in the formation of granulomas has been widely studied ⁹⁷⁻⁹⁹. Biallelic polymorphisms at position -308 in the

promoter region of the *TNFA* gene and in the first intron of the *LTA* gene have been associated with variation in TNF α production and with susceptibility to infectious and autoimmune diseases⁹⁸. Association studies on genetic polymorphisms in promoter region have raised the possibility that this cytokine might influence the predisposition or severity of sarcoidosis^{70, 98}. Except for a higher allele frequency of a rare *TNFA* allele found in patients suffering from acute sarcoidosis, other studies have not found significant association of either polymorphism with sarcoidosis^{98, 100}. Among the chemokines, many association studies have focused on gene polymorphisms in CC chemokine receptor 2 (*CCR2*) and CC chemokine receptor 5 (*CCR5*) because the products of these genes are expressed on the surface of CD4+ T cells. The *CCR2* valine to isoleucine substitution at amino acid position 64 appears to have a protective effect¹⁰¹, whereas the 32 bp deletion in the promoter of the *CCR5* gene increases susceptibility to sarcoidosis¹⁰². A haplotype study indicated that variations in the *CCR2* gene are associated with different sarcoidosis phenotypes¹⁰³.

Clinical similarities between Blau syndrome, Crohn's disease and sarcoidosis suggest the presence of shared susceptibility genes. Linkage on chromosome 16 has been established for both Blau syndrome and Crohn's disease¹⁰⁴⁻¹⁰⁷. Later, several independent association studies have demonstrated that distinct alleles of the *CARD15* gene contribute to susceptibility for Crohn's disease¹⁰⁸⁻¹¹⁰. The *CARD15* gene resides on the long arm of chromosome 16 (16q12). The protein is composed of two caspase recruitment domains (CARD), a nucleotide oligomerisation domain (NOD) and a stretch of leucine rich repeats (LRR). The putative function of the *CARD15* protein also contributes to the hypothesis of shared susceptibility genes. *CARD15* may be involved in activation of the nuclear regulatory factor kappa B (NF κ B) pathway, which is activated after stimulation with microbial lipopolysaccharides. High NF κ B activity is characteristic feature of both disorders^{111, 112}. Furthermore, mutations in the *CARD15* gene could influence the regulation of immune cell apoptosis¹¹³.

Blau syndrome mutations were found in the NOD domain of the *CARD15*, while mutations associated with Crohn's disease are located in the LRR region. Thus, alterations in different regions of the gene could cause or contribute to different granulomatous disorders.

1.3. Aims of the study

1. Identification of HLA-independent susceptibility marker for sarcoidosis in the chromosome 6 linkage region.
2. Evaluation of *CARD15* as a candidate gene for sarcoidosis.

2. Materials and methods

2.1. Materials and equipment

Table 1. Part 1: materials

Material	Manufacturer/Supplier, Country
100bp DNA ladder	Invitrogen, Karlsruhe, Germany
2.2 ml storage plate (96 well)	ABgene, Epsom, UK
384 deep well storage plate	ABgene, Epsom, UK
Agarose	Eurogentec; Köln, Germany
Taq® DNA polymerase	Qiagen, Hilden, Germany
AmpliTaq® Gold DNA polymerase	Applied Biosystems; Darmstadt, Germany
Bacilol®	Bode Chemie, Hamburg, Germany
BigDye Terminator Ready reaction kit	Applied Biosystems, Darmstadt, Germany
Bromphenol blue	Sigma, München, Germany
dNTP set (100mM solutions- each 100µM)	Amersham Biosciences; Freiburg, Germany
Ethanol	Merk; Darmstadt, Germany
Ethidium bromide solution (10mg/ml)	Invitrogen; Karlsruhe, Germany
GeenAmp PCR buffer system (10x buffer w/o MgCl ₂ ; 25mM MgCl ₂)	Applied Biosystems, Darmstadt, Germany
Invisorb Blood Giga Kit	Invitex; Berlin, Germany
MgCl ₂	Merck; Darmstadt, Germany
MicroAmp®optical 96-well reaction plate	Applied Biosystems, Darmstadt, Germany
MicroAmp®single strips	Applied Biosystems, Darmstadt, Germany
MicroAmp® single tubes	Applied Biosystems, Darmstadt, Germany
Microtiter 384-well plates	Sarstedt; Nürnberg, Germany
Microtiter 384-well plates	Greiner Bio-One GmbH, Frickenhausen, Germany
Microtiter 96-well plates	Sarstedt; Nürnberg, Germany
Microtiter 96-well plates	Costar Corning Incorporated; Cambridge, MA, USA
Microtiter plates, 96 well, round bottom with lid	Sarstedt; Nürnberg, Germany

Table 1. Part 2: materials

Material	Manufacturer/Supplier, Country
PicoGreen®	Molecular Probes Europe BV; Leiden, The Netherlands
Pipette (serological, sterile with filter 5/10/25ml)	Sarstedt; Nümbrecht, Germany
Pipette tips with filter (10/200/1,000µl)	Sarstedt; Nürnberg, Germany
SAP shrimp alkaline phosphatase	Amersham Biosciences; Freiburg, Germany
Sephadex powder (G50 superfine)	Amersham Biosciences; Freiburg, Germany
Sepahdex spin column plates MAHVN 4550	Amersham Biosciences; Freiburg, Germany
SmartLadder DNA marker	Eurogentec; Köln, Germany
TAE Buffer 25x ready pack	Amresco; Solon, OH, USA
TaqMan® Universal PCR Master Mix	Applied Biosystems; Darmstadt, Germany
TBE Buffer 10x ready pack	Amresco; Solon, OH, USA
Tubes (0.5 / 1.5 / 2.0 ml)	Eppendorf; Köln, Germany
Tubes (0.5 / 1.5 / 2.0 ml)	Eppendorf; Köln, Germany
Tubes, flat bottom (60 ml)	Sarstedt; Nümbrecht, Germany
Tubes, sterile (15 ml)	Sarstedt; Nümbrecht, Germany
Tubes, sterile (50 ml)	BD Biosciences; Heidelberg, Germany
Xylene Cyanol FF	Sigma; München, Germany

Table 2. Equipment

Equipment	Manufacturer/Supplier, Country
10/100/300/1000 μ l single/multi-channel pipettes	Eppendorf; Hamburg, Germany
7900HT Sequence Detection Systems	Applied Biosystems; Darmstadt, Germany
ABI PRISM®7700 Sequence Detector	Applied Biosystems; Darmstadt, Germany
Biometra® T Gradient	Whatman Biometra GmbH; Göttingen, Germany
Biometra® T1 Thermocycler	Whatman Biometra GmbH; Göttingen, Germany
GeneAmp®PCR system 9700	Applied Biosystems; Darmstadt, Germany
Gibco BRL Electrophoresis Power Supply™ 250 EX	BioRad; München, Germany
Gibco BRL Horizontal Gel Electrophoresis Apparatus™	BioRad; München, Germany
Heraeus Kelvitron®t	Kendro; Hanau, Germany
Heraeus Labofuge 400	Kendro, Hanau, Germany
High Performance Ultraviolet Transilluminator	VWR; Hamburg, Germany
Horizontal Electrophoresis Apparatus	BioRad; München, Germany
Hydra™384 Robbins Scientific®	Dunn Labortechnik GmbH; Asbah, Germany
Hydra™96 Robbins Scientific®	Dunn Labortechnik GmbH; Asbah, Germany
Platesealer ALPS-300	Abgene, Epsom, UK
Power Pack 300 Electrophoresis Power Supply	BioRad; München, Germany
TECAN Genesis RSP 150	Tecan Deutschland GmbH; Crailsheim, Germany
TECAN GENESIS Workstation 150	Tecan Deutschland GmbH; Crailsheim, Germany
TECAN GENESIS Workstation 200	Tecan Deutschland GmbH; Crailsheim, Germany
TECAN Spectrafluor Plus	Tecan Deutschland GmbH; Crailsheim, Germany
Vortex-GENIE 2® G-560E	Scientific Industries, Inc., Bohemia, N.Y., USA

2.2. Electronic data bases

Applied Biosystems	http://www.appliedbiosystems.com/
BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
Celera Discovery Systems	http://www.celera.com/
dbSNP	http://www.ncbi.nlm.nih.gov/SNP/index.html
Gene Cards	http://bioinfo.weizmann.ac.il/cards/
HLA database	http://www.ebi.ac.uk/imgt/hla/align.html
Locus Link	http://www.ncbi.nlm.nih.gov/LocusLink/
MAP viewer	http://www.ncbi.nlm.nih.gov/mapview/
NCBI	http://www.ncbi.nlm.nih.gov/
PubMed	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed
Repeat Masker	http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker
UCSC Genome Browser	http://genome.ucsc.edu/

2.3. Participants

2.3.1. Sarcoidosis population for the chromosome 6 association study

Sarcoidosis patients and close relatives were contacted between years 1996 and 2003 through specialised hospitals and general practitioners, or through the German sarcoidosis patient organisation (Deutsche Sarkoidose-Vereinigung e. V., www.sarkoidose.de). Others were enrolled after calls for participation in the study that were forwarded by health insurance institutions. Diagnoses were made on the basis of the International Consensus Statement on Sarcoidosis^{33, 114}. Patients from sarcoidosis families were interviewed by telephone about their individual and familial sarcoidosis history. All patients (families and single affected patients) completed a questionnaire about the course of their disease. Physicians were contacted to confirm the diagnosis, and the presence of sarcoidosis was additionally verified by biopsy in 78% of cases from the “exploratory” sample, 90% of patients from the “extension” sample, and in 88% of patients from the “replication” sample. For the remainder, the clinical course and the radiology and laboratory data were regarded as sufficiently consistent with the diagnosis of sarcoidosis¹¹⁴.

Observed sarcoidosis phenotypes were grouped into ten classes in the entire sample presented in this study (for detailed sarcoidosis phenotype classification and description see supplement information 9.1.2.). 31% of all affected individuals had an acute form of sarcoidosis (mainly with Löfgren syndrome) and 61% were classified as chronic patients. Both acute and chronic patients had involvement of the lung.

All patients were of German origin. All participants gave written informed consent for participation in the study. The protocols were approved in writing by the institutional ethics and data protection authorities. An overview of the patient samples is given in table 3.

Table 3. *Sarcoïdosis sample for the chromosome 6 association mapping study.* Upper part: non-overlapping categories of sampling units (families, trios, singleton cases); lower part: summary of case/patient numbers. The total number of independent cases equals the sum of the sampling units in the upper part. The experiments performed in different combinations of patient samples are depicted in the lower part of the table.

Sampling unit	Patient sample			
	exploratory	extension	replication	Full
Affected sib pairs				
(ASP) – 2 sibs	70	9	10	89
ASP – 3 sibs	8	–	–	8
Complex families	11	5	–	16
Trios (complete)	153	72	203	428
Trios (one parent missing)	10	79	62	151
Single patients	13	65	177	255
Total number of independent cases	265	230	452	947
Total number of patients (male/female)	372 (163/209)	248 (123/125)	462 (175/287)	1082 (461/621)
Average age at diagnosis (SD) in years	34.7 (10.3)	37.2 (11.4)	37.3 (11.3)	36.4 (11.1)
Experiment	Initial SNP screen			Differentiation of <i>DRB1</i> - <i>BTNL2</i> effects
	SNP fine mapping		Replication	

2.3.2. Control group individuals

A group of 517 age- and sex-matched healthy blood donors of German origin were included as control group in the association analysis.

2.3.3. Sequencing patients for association mapping experiment on chromosome 6

Only for a minority of sequence variations annotated in databases, SNP allele frequency is indicated. Even more, part of annotated SNPs derive from mismatches by comparing sequences of overlapping clones. Such mismatches can arise due to mutations introduced in the clone sequence so that the variation in reality does not exist as an SNP. In regions where no SNPs in databases are annotated but for a higher density of a map more SNPs are needed, variations can be detected through resequencing of genomic DNA. For all these reasons, resequencing of genomic DNA of 47 unrelated individuals was used. To not overlook specific sequence variations that could be the disease causing mutations and thus have a very low frequency in sample of healthy individuals, sarcoidosis patients have been chosen.

2.3.4. CARD15 gene analysis samples

The *CARD15* gene was investigated in the following two sarcoidosis populations: i) families with two or more members suffering from sarcoidosis (henceforth referred to as "sarcoidosis families"), ii) unrelated sarcoidosis patients without a family history of sarcoidosis, together with their healthy parents (henceforth referred to as „sarcoidosis trios”). For the case-control analysis, a sample of healthy control individuals was used. The complete families panel comprised 302 patients (135 males, 137 females; average age at inclusion (mean±SD) 51±13.1 yrs, average age at diagnosis 35±11.3 yrs, range 12-73 yrs) and 153 close relatives without sarcoidosis (64 males, 89 females). The trios sample consisted of 127 patients (49 males, 78 males; average age at inclusion 34±6.8 yrs, average age at diagnosis 33±6.9 yrs, range 11-51 yrs) and 254 parents without sarcoidosis. The control groups comprised 265 healthy German Caucasian blood donors recruited through the Dept of Transfusion Medicine at the Kiel University Hospital, Kiel, Germany (118 males, 147 females: average age at inclusion 51±8,2 yrs).

The sample of 138 German families of Caucasian origin included 66 families with two or three affected sibling pairs, 41 families with affected parent-offspring pairs, 19 families with two more distantly related patients and 12 families with a more complex structure involving more than two patients. The sarcoidosis trio sample comprised 127 spontaneous cases with unaffected parents. (see the detailed *CARD15* study family sample information in 9.1.3.). All

CARD15 study patients were a subset of the chromosome 6 association mapping study sample.

2.4. DNA samples

2.4.1. Isolation of genomic DNA from blood

Genomic DNA was extracted using a slightly optimised Invisorb®Blood Giga Kit for DNA isolation (Invitek, Germany). 10ml of whole blood were incubated with 30ml of cold Buffer I for 10 minutes at a room temperature to achieve lysis of erythrocytes and centrifuged at 3000 rpm for 3 minutes. After the supernatant was carefully removed, the procedure was repeated with 20ml of Buffer I to achieve a better purification of leucocytes from erythrocytes. The cells were centrifuged at 3000 rpm for 3 minutes. The cell pellet containing leucocytes should be white, if this was not the case then the second step was repeated. The pellet was then resuspended in 3ml of Buffer II and 50µl of Proteinase K and incubated overnight in a 60°C water bath with continuous shaking. During this step, the reagents lyse the leucocytes and the DNA becomes free from nuclei. For protein lyses, 1,8ml of Buffer III was added, the mix was vigorously shaken and incubated on ice for 5 minutes. After incubation, the mix was centrifuged for 15 minutes at 5000rpm. The cleared supernatant was transferred into a 15ml centrifuge tube. For the precipitation of the genomic DNA, 10ml of 96% Ethanol was added and the tube was carefully inverted several times. The precipitated DNA was collected with a pipette tip and transferred into 2ml eppendorf tubes containing 1 ml of 70% ethanol. The DNA pellet was rinsed with ethanol, vortexed and then centrifuged at 13000rpm for 2 minutes. In cases where no DNA pellet was present, patients were contacted for an additional blood sample. After the rinsing step, the ethanol was removed by pipette and samples were dried for 10 minutes at a room temperature. The purified genomic DNA was resuspended in 500µl of 1xTE buffer and stored at +4°C for a short time or at -20°C for longer periods.

2.4.2. Plates for genotyping with TaqMan assays

The DNA samples for di-allelic genotyping were arrayed into 96- and 384-well format plates. Every 96 well plate contained not more than 92 DNA samples leaving the rest of space for controls. Complete families were included in the same plate. There were at least two wells with CEPH (Centre d'Etude du Polymorphisme Humain, Paris, France) cell line samples in

each plate as positive amplification controls. One to four wells were left empty for a negative amplification control. Every 96 well plate received a unique identifier and was entered into the laboratory data base system. 384 well format plates were designed by arraying four 96 well plates. A barcode system has been implemented for the identification of each individual plate. Samples were pipetted into 96 and 384 deep-well plates (Abgene, Epsom, UK) and the DNA concentration adjusted through two steps of dilution to 10ng/ μ and 0,5ng/ μ l using a TECAN Genesis RSP 150 multipipetting robot (Tecan Deutschland GmbH; Crailsheim, Germany). The aliquots of DNA (5 μ l of 0,5ng/ μ l DNA-TE aliquot) were distributed via either a 96- or 384-channel Robbins Scientific Hydra microdispenser into the 96-well and 384-well plates (Costar Corning Incorporated; Cambridge MA, USA, Greiner Bio-one GmbH; Frickenhausen, Germany) and dried at 60°C (~2 hours). Dry DNA plates were covered with Plate Sealer (AB Gene, Hamburg, Germany) and stored at room temperature until use.

2.4.3. HLA-DRB1 typing plate design

Standard 96-well format dry DNA plates with sarcoidosis patients and control sample DNA (as described in the chapter above) were used for genotyping the multiallelic marker *HLA-DRB1*. Additional DNA samples with known *HLA-DRB1* genotypes were used as positive controls for typing of *HLA* alleles (see table 4).

1 μ l of the 2.5ng/ μ l positive control DNA was pipetted into two empty (no-template) positions on each 96-well plate. Amplification of a positive DNA sample precluded false interpretation of amplification results: i) if both positive controls were successfully amplified but no patient was amplified, then the plate was still recorded as working; ii) if positive amplification was achieved in patients but not in both controls then the plate was discarded, since the result could be obtained from mixed-up primers. Allele-groups 6 and 7 had ambiguous positive DNAs (DRB1*03,*09 for amplification group 6 and DRB1*03,*10 for amplification group 7) because the alleles DRB1*09 and *10 are rare there was little likelihood that homozygotes would be observed. Results of these two amplification groups were accepted based on a comparison of the number of positive amplifications in the plate and the expected population allele frequency. Contamination with DNA carrying the frequent HLA-DRB1*03 allele or usage of amplification group 3 primers by mistake, would have given up to a 20 times higher frequency of positive amplification. Consistency of CEPH (positive controls for TaqMan

assays) final genotypes in all plates was checked after all allele-groups were amplified and final genotypes were defined.

The remaining no-template wells served as controls of negative amplification for the PCR reagents.

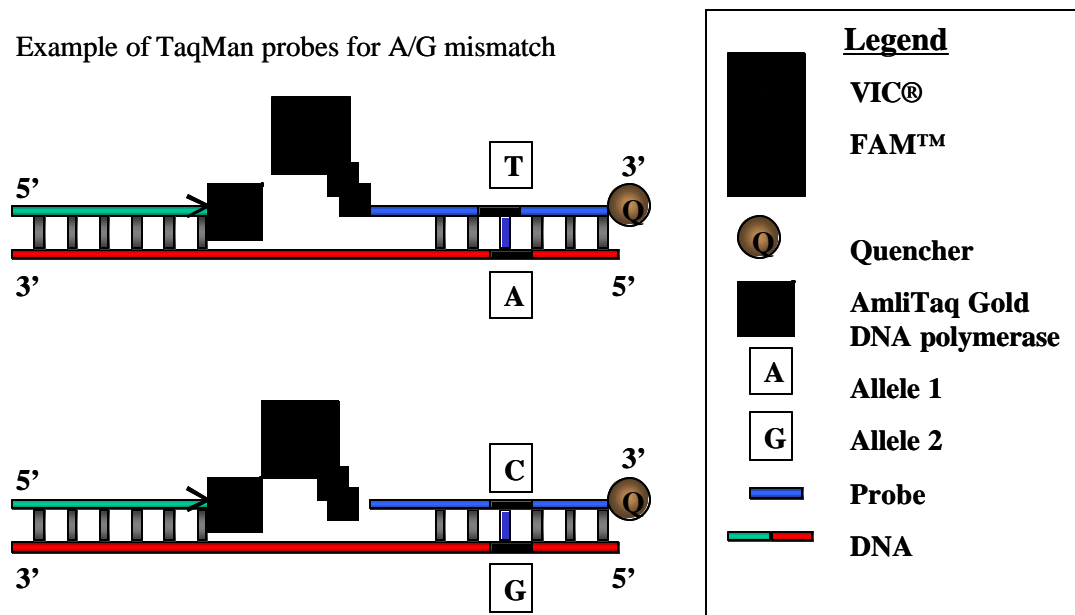
Table 4. Known HLA-DRB1 genotypes of patients used as positive controls for amplification of HLA-DRB1 allele-groups 1 to 7. Positive DNA: patient DNA with known genotype used as positive control for respective HLA-DRB1 allele-group.

<i>HLA-DRB1</i> allele amplification group	<i>HLA-DRB1</i> alleles amplified by corresponding allele group	Positive DNA genotypes
1	DRB1*01	DRB1*01
2	DRB1*15,*16	DRB1*15,*16
3	DRB1 *03,*08,*11(not *1122/30),*12,*13,*14 (not *1410/39)	DRB1*08,*12
4	DRB1*04,*1122,*1410	DRB1*04
5	DRB1*07	DRB1*07
6	DRB1*09	DRB1*03,*09
7	DRB1*10	DRB1*03,*10

2.5. Di-allelic genotyping

To determine the genotype at known SNPs, di-allelic genotyping was performed using a TaqMan 5' nuclease fluorogenic assay¹¹⁵. An oligonucleotide probe that is specific for the target (SNP) to be amplified is labeled with the fluorescent reporter dyes FAM™, VIC® or TET at the 5' end and a quenching molecule TAMRA at the 3' end. During the extension step of a PCR the AmpliTaq Gold® polymerase digests the probe bound to the target and separates the fluorescent tag from its quencher molecule thus permitting fluorescence. Consequently, the fluorescence signal(s) generated by PCR amplification indicate(s) the alleles that are present in the sample (see fig.4).

-Fig. 4. Schematic view of TaqMan® assay.



Fluorescence results of diallelic genotyping were visualized using laser scanning technology (ABI Prism® 7700 Sequence Detection System for 96-well format plates or ABI Prism®7900HT Sequence Detection System for 384-well format plates). The laser light is targeted into 96 or 384 wells via a multiplexed array of optical fibres. The fluorescent emission returns via the fibres and is directed to a spectrograph with a charge-coupled device camera. Possible results of allelic discrimination are as following: i) TET (VIC® for TaqMan®-MGB assays) fluorescence only for allele 1 homozygosity, ii) FAM™ fluorescence only for allele 2 homozygosity, ii) both fluorescent signals for heterozygosity.

Diallelic genotyping of individuals plated on 96-well and 384-well plates was performed using TaqMan® and TaqMan®-MGB assays. The TaqMan®-MGB probe¹¹⁶ is an alternative probe for TaqMan® (demonstrated in fig.4). These probes differ in the following features: i) there is a non-fluorescent quencher attached to the 3'end, which allows the fluorescence of the reporter dye to be measured more precisely against a lower background fluorescence, ii) the 3' end has a minor groove binder (MGB), which increases the melting temperature of the probes and thus permits shorter probes. Short probes exhibit greater differences in annealing temperature values between matched and mismatched probes. That feature provides a higher quality of allelic discrimination.

2.5.1. Di-allelic assay design and genotyping

TaqMan® and TaqMan®-MGB assays were utilised for the genotyping of SNPs in a panel of sarcoidosis and control DNAs. Some pre-optimised assays for allelic discrimination were available from the ABI company (Assay-on-Demand, Assay-by-Design, Applied Biosystems Inc., Foster City, CA, USA) and from laboratory resources built up for other projects^{117, 118}. Several assays in more complicated sequence regions were self-designed by hand as TaqMan®-MGB assays.

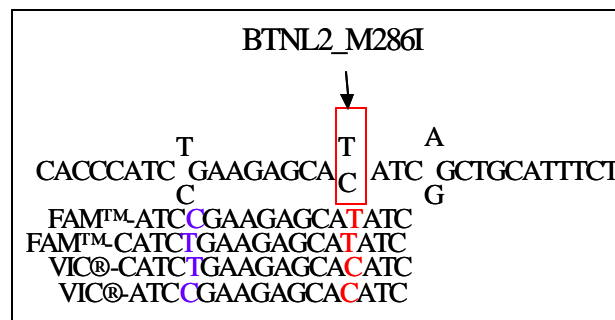
To design a TaqMan®-MGB assay the genomic region of interest has to be analysed in detail. It is essential that the target region be first analysed for repetitive elements (<http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker>). Probes can only be designed for a unique sequence with one SNP. The sequence should contain no repetitive elements or additional SNPs. Non-repetitive sequence is not obligatory, but is certainly desirable for primers. The presence of other SNPs is not permitted at the primer binding sites. If the sequence had many repetitive elements and the positioning of both primers on unique sequence was not possible, at least one of the primers was positioned on non-repetitive sequence (for detailed sequencing and mutation detection description, see chapter 2.7.).

TaqMan®-MGB diallelic assays were designed using the software Primer Express 2.0 (Applied Biosystems Inc., Foster City, CA, USA). Probes and primers were designed according to the criteria of the supplier. Accordingly, the melting temperature of the probe must be 7°C higher than the primer melting temperature and range between 65°C and 67°C. The melting temperatures of both probes have to be as close as possible with a maximum difference of 1°C. The C and G nucleotide content should be 20-80%. The probe must be 13 to 20 base pairs in length and contain less than four contiguous G nucleotides. The total number of G nucleotides should not exceed the number of C nucleotides in the probe. The probe must not have a G at the 5' end. The polymorphism has to be located in the interval between the middle and 3' end of the probe, but not at the last 2 nucleotides. The primers were designed to match the probes. The primer sequence was selected to as to meet the criterion of a 7°C lower T_m difference than the probe. The T_m interval therefore was between 58°C and 60°C with a less than 2°C difference between the two primers. The CG content ranged between 30% and 80% with a maximum of 2 out of 5 C nucleotides at the 3' end. Primers with high nucleotide variation at the 3'end were preferred. Primer length varied

between 9 and 40 base pairs. The amplicon length varied between 50 and 150bp. The 3' end of the forward primer (located on the same strand as the probe) had to be positioned as close to the probe as possible, but without overlapping. As the reverse primer only amplifies the target sequence, this may overlap with the probe (but not with the SNP). If the assay design did not meet the described conditions the complement strand was examined to facilitate the best choice of assay.

A special design was applied to the measurement of sequence variation at the BTNL2_M286I site. The target SNP was flanked by another two SNPs and hence a standard probe could not be designed. Therefore, a special four probe combination was chosen. For each allele at the target SNP two FAM and two VIC labelled probes were designed such that each allele at an adjacent SNP would be targeted by both FAM and VIC probes and have no impact on the discrimination at the target SNP (see fig.5).

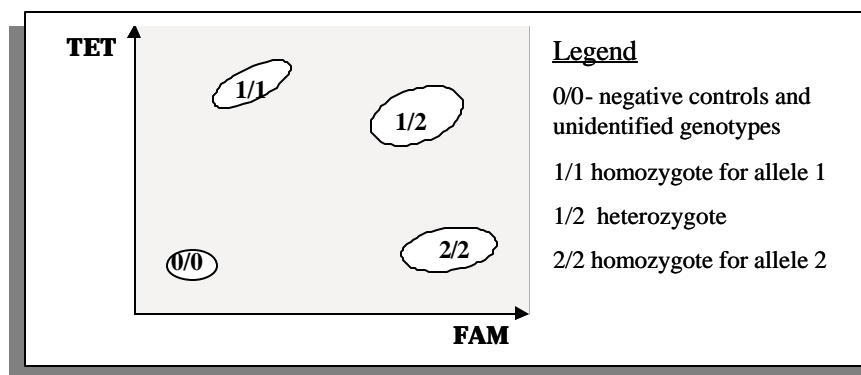
Fig.5. TaqMan® assay using four-probe combination. Two probes labelled with FAM™ had a T nucleotide at the BTNL2_M286I variation site, but both C and T nucleotides at another variation site located eight nucleotides upstream. The same was applied for the probes labelled with VIC® fluorescent dye.



Self-designed TaqMan®-MGB assays were optimised for annealing temperature. The experiment was performed on 96-well plates with 2.5ng control sample DNA. Two annealing temperatures were tested: 60°C and 62°C (for standard cycling protocol, see next page). Supplied probes and primers were adjusted to 100µM storage concentration with double distilled water. Working dilutions were 10µM for probes and 20µM for primers. The final reaction mixture volume was 5µl and contained 200nmol of each probe and 900nmol of each primer. 2.5µl of 2-times TaqMan® Universal MasterMix (Applied Biosystems Inc., Foster City, CA, USA) and DDW were added into reaction up to the 5µl final volume. The Thermocycler GeneAmp®PCR System 9700 (Applied Biosystems Inc., Foster City, CA, USA)

was employed for TaqMan®-MGB assay optimisation. Standard cycling conditions with alternative annealing temperatures were applied. After cycling, plates were visualised in the ABI Prism®7700 Sequence Detector (Applied Biosystems Inc., Foster City, CA, USA). The assay had to meet the following quality criteria. For diallelic assays, only a maximum of 3 clouds of dots corresponding to the 3 different genotypes (two homozygotes and one heterozygote) was satisfactory. One additional cloud located at the cross of the x- and y-axis had to contain only no-template control samples (represented as genotype “0/0” - see fig.6). The observed genotypes of the positive controls (the CEPH cell line DNA) had to be the same across all plates genotyped for that assay. Missing genotype data was not allowed to exceed 5% of the sample. The genotype frequency distributions in the independent control samples had to satisfy Hardy-Weinberg Equilibrium.

Fig.6. Schematic representation of fluorescent measurement of TaqMan® assay. The assay measures fluorescence of biallelic variations. A maximum of three genotype clouds and one negative control cloud are detectable in the view.



96 and 384 platforms were employed for the genotyping of all diallelic SNPs in the sarcoidosis sample and the healthy controls. The PCR reaction mastermix was prepared for at least one plate. For the 96-well reactions, primers and probes were mixed with the 2.5µl 2 times concentrated TaqMan® Universal PCR Master Mix (Applied Biosystems; Darmstadt, Germany) and diluted with DDW to achieve a 5µl reaction volume. Optimal final concentrations for primers and probes in all assays of this study varied from 50nmol to 900nmol and from 100nmol to 200nmol respectively (see supplement 9.1.5.). The mix was applied to the plates containing 2.5ng dried DNA. The amplification step was performed in the GeneAmp®PCR system 9700 (Applied Biosystems Inc., Foster City, CA, USA). The thermal cycle had the following profile: a 2min hold step for optimal AmpErase uracil-

N.glycosylase activity followed by a 10min hold step at 95°C for AmpliTaq Gold DNA Polymerase activation (both are components of the TaqMan® Universal PCR Master Mix). This was followed by 45 PCR cycles consisting of a denaturation step at 95°C for 15sec and then a 1min annealing step at the optimal temperature (which ranges between 58°C and 64°C -see supplement 9.1.5). The plates were read in the 7900HT Sequence Detection System (Applied Biosystems; Darmstadt, Germany). The resulting fluorescence was analysed using an integrated software to manually call the genotypes on each plate. Plates were repeated or the assay was excluded from analysis, if the performance rate was lower than 95%. Other quality criteria were the same as applied for the assay optimisation.

For higher automation, a 384-well platform was employed. Reaction mix was applied to the plates with the aid of The TECAN GENESIS Workstation 150 (Tecan Deutschland GmbH; Crailsheim, Germany) and thermal cycling was performed in Biometra® T1 Termocyclers (Whatman Biometra GmbH; Göttingen, Germany). Fluorescence was evaluated in the ABI PRISM®7700 Sequence Detector (Applied Biosystems; Darmstadt, Germany).

The resulting genotype scoring was integrated into an internal data base and linked with the patient information (i.e., pedigree information, patient identification number, gender and affection status).

2.5.2. SNP selection for chromosome 6 mapping

The interval for association mapping on chromosome 6 was selected based on the linkage result in the genome scan⁵⁸. Defined by a decrease in LOD score by 1.5 units or less, this interval represented an approximate 99% confidence region for linkage. 69 biallelic markers for the initial screening of the region were selected largely from the „peri-HLA SNP set” that was previously evaluated for association studies on chromosome 6¹¹⁷. The SNP set covered the target region located at 27.80 – 44.14 Mb (NCBI release 32) on chromosome 6.

2.5.3. Typing of the CARD15 gene

In total, 1101 individuals were genotyped for the three Crohn’s disease associated *CARD15* polymorphisms: SNP8 (2104C>T; R702W; nomenclature according to ref.¹¹⁹), SNP12 (2722G>C; G908R), SNP13 (3020insC; 1007fs) and SNP5 (802C>T; P268S) were selected as

the susceptibility haplotype tagging SNP. Furthermore, SNP5 is located in exon 4 of the *CARD15* gene, the exon where the mutations causing Blau syndrome reside. The TaqMan™ system was used for the genotyping. In this thesis, the rarer alleles of these polymorphisms are annotated as allele 1. For detailed primer and probe information see supplement 9.1.4., and for genotyping conditions supplement 9.1.5.

2.6. Multi-allelic genotyping

2.6.1. *The Problem of multi-allelic genotyping*

Current HLA genotyping methodologies include group-specific PCR¹²⁰, genomic or cDNA sequencing¹²¹, and hybridisation¹²² methods, employed either individually or in combination. The standard methods of SNP genotyping for non-HLA loci, for example the TaqMan™ system, face significant problems in highly polymorphic regions such as the informative HLA molecule exon sequences. There is an extremely high density of SNPs in the HLA region, which makes probe and primer design very difficult. This high diversity is further complicated by a complex haplotype structure that makes successful allele resolution, even with multiple probes, impossible. The standard techniques require a universal and stable underlying PCR. For many HLA loci, the design of such primers is impossible, because either undesirable co-amplification of pseudogenes or lack of amplification of the relevant alleles at the locus of interest occurs.

The DRB1 locus has all these – from a genotyping point of view – undesirable characteristics.^{123, 124} (HLA Database at www.ebi.ac.uk/imgt/hla). Therefore, an appropriate method that genotype this multi-allelic locus had to be employed.

2.6.2. *Genotyping of HLA-DRB1 by allele-group PCR*

High resolution typing of specific *HLA-DRB1* alleles was not required in our situation, because we were interested in tagging of “risk haplotypes” and differentiation of the genetic effect from the *BTNL2* gene. Therefore, a system of non-overlapping allele groups was exploited in order to amplify the basic alleles of the *DRB1* gene.

Primers for the detection of the seven major *DRB1* allele groups were adopted from published reports ^{125, 126} and are listed in table 5. Specific *HLA-DRB1* alleles were merged into allele groups: DRB1*01 (allele-group 1); DRB1*15,*16 (2); DRB1 *03,*08,*11(not *1122/30),*12,*13,*14 (not *1410/39) (3), DRB1*04,*1122,*1410 (4), DRB1*07 (5), DRB1*09 (6) and DRB1*10 (7) (see table 5).

Table 5. Primers used for the amplification of *HLA-DRB1* allele groups. Allele group-specific primers were adapted according to ref ^{127, 128}. Specific seven forward primers and one common reverse primer 5'-CCGCTGCACTGTGAAGCTCT-3' were used for all amplifications ¹²⁹. Primers were selected for use in a Caucasian sample. Certain rare DRB1 alleles will not be amplified with these primers, but their combined frequencies are below 0.05% and thus not relevant for the application needed in this context.

Amplification group	Primer	<i>DRB1</i> alleles detected
1	5'-TTCTTGTGGSAGCTTAAGTT-3'	DRB1*01
2	5'-TTCCTGTGGCAGCCTAAGAGG-3'	DRB1*15, *16
3	5'-CACGTTTCTTGGAGTACTCTAB-3'	DRB1*03, *08, *11(not *1122/30), *12, *13, *14 (not *1410/39)
4	5'-GTTTCTTGGAGCAGGTTAAAC-3'	DRB1*04, *1122, *1410
5	5'-CCTGTGGCAGGGTAARTATA-3'	DRB1*07
6	5'-CCCAACCACGTTTCTTGA-3'	DRB1*09
7	5'-AGACCACGTTTCTTGGAGG-3'	DRB1*10

To achieve optimal PCR conditions, MgCl₂ (1,0;1,5;2,0;2,5;3,0;3,5;4,0μl) and Taq polymerase (0,02;0,03;0,06;0,09;0,12μl) concentrations were titrated for all seven PCR reactions. PCR optimisation was performed in a 25μl reaction mixture. In each case, optimal PCR conditions were established using a panel of human DNAs of known genotype taken from the German HLA quality control panel. These DNAs were used in all allele -group amplification experiments as a quality controls for positive amplification (for control DNA panel see table 4).

All allele-group genotyping reactions were carried out in 10μl reaction mixtures with 1μl(10x) buffer, 0.2mM dNTPs (each), 0.2μM primer (each) and the MgCl₂ and Taq

polymerase concentrations listed in table 6. All reactions employed AmpliTaq Gold® with Gene Amp, Applied Biosystems, USA.

Table 6 .PCR conditions used for DRB1 allele-group amplification. Optimal MgCl₂ and Taq DNA polymerase concentrations varied between amplification groups.

Amplification group	MgCl ₂ mM	Taq DNA polymerase U/μl
1	2	0.024
2	2	0.030
3	3.5	0.036
4	2	0.024
5	2	0.018
6	3	0.036
7	2	0.036

Touchdown PCR programs were used for all primer pairs (performed with the GeneAmp PCR System 9700, Applied Biosystems, USA). Optimal annealing temperatures for each primer pair were identified via a temperature gradient (employing a Biometra® T Gradient, Whatman Biometra GmbH; Göttingen, Germany). Optimal annealing temperatures was identified through inspection of the intensity of the amplicon bands when visualised on 1.5% agarose gel (Agarose, Eurogentec; Köln, Germany). Electrophoresis conditions were 30min at 110V. The PCR cycle conditions used for the genotyping of the entire sarcoidosis sample and controls are listed in table 7.

Table 7. PCR temperature profiles used for DRB1 allele-group amplification.

Allele-amplification group				Time	Number of cycles	Process
7	1/3	6	2/4/5			
95°C	95°C	95°C	95°C	12 min		Hot start
95°C	95°C	95°C	95°C	30 sec	16	Denaturation
56°C	61°C	63°C	65°C	30 sec (touchdown - 0.5°C)		Annealing
72°C	72°C	72°C	72°C	30 sec		Extention
95°C	95°C	95°C	95°C	30 sec	19	Denaturation
48°C	53°C	55°C	57°C	30 sec		Annealing
72°C	72°C	72°C	72°C	30 sec		Extention
10°C	10°C	10°C	10°C	Pause		Hold

2.6.3. Development of a non-electrophoretic method for high-throughput DRB1 allele-group identification

The majority of methods for the visualisation of *HLA* allele amplification products employ an electrophoretic approach. However, this technique does not scale well in a high throughput setting. Thus, for the characterisation of PCR products a non-electrophoretic method was adapted. The method was previously described for genotyping a G/A transition in the p53 proto-oncogene of Chinook salmon¹²⁵. The methodology is based on ability of the fluorescent dye PicoGreen® (Molecular Probes Europe BV; Leiden, The Netherlands) to bind specifically to dsDNA and thus to identify amplified product by its increased fluorescence.

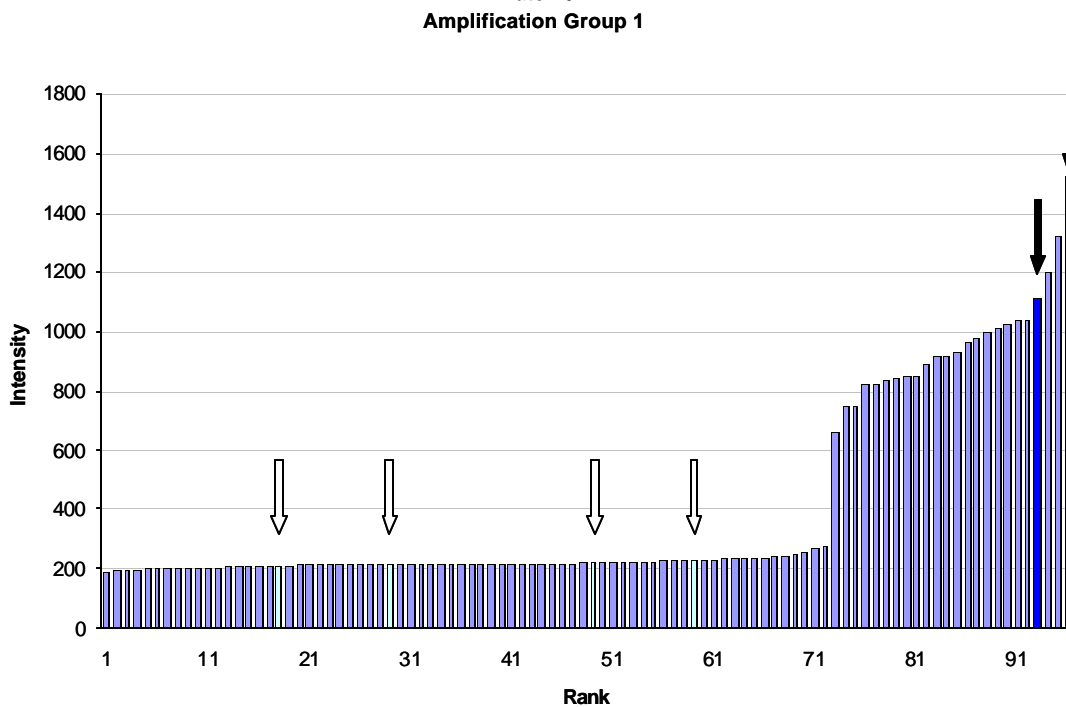
Initially, the validity of the non-electrophoretic measurement assay was tested by comparing PicoGreen® measurement results with those obtained by agarose gel assay. One 96-well format plate with ninety individuals was genotyped for all seven allele groups. 5µl of the PCR products were visualised on an agarose gel stained with ethidium bromide (1.5% agarose gel (Agarose, Eurogentec; Köln, Germany); electrophoresis conditions were 30min at 110V). Another 5µl of PCR products were incubated with PicoGreen® and the fluorescence was

measured. Results of both methods coincided for all ninety samples through all seven amplification groups.

Five microliters of PCR product were transferred into 96 well plates (Microtiter plates, 96 well, round bottom with lid, Sarstedt; Nürnberg, Germany) containing 45µl of PicoGreen® diluted 1:400 with 1xTE buffer. The plates were vortexed, incubated for 5 min, and then read in a TECAN Spectrafluor plus Fluorometer (TECAN, Männedorf, Switzerland) with excitation wavelength 485nm. The raw data files were exported for each plate separately.

PCR plates were read in a TECAN Spectrafluor plus (Tecan Deutschland GmbH; Crailsheim, Germany), which generated “device-specific artificial units”. Analysis of raw data was performed in Microsoft Excel®. Measurements were automatically arranged based on signal intensity. A clear, sudden increase in intensity in the „positive” samples (samples that contain the amplified allele) corresponded to the dsDNA of a PCR product (see fig. 7).

Fig.7. Fluorescence reading of a 96-well plate for the amplification group 1. The y-axis denotes the signal intensity of the fluorescence measurement of the TECAN Spectrafluor plus in device-specific artificial units. Based on the signal intensity, measurements 1 to 96 of a 96-well plate are ordered on the x-axis. The positions of the positive and negative controls on the graphic are denoted with solid and open arrows, respectively.



In this example, samples that had a fluorescence more than 600 units were assigned as „positives” ,i.e., containing that allele. The „negative” samples contained only original DNA that was not amplified, because the allele was not present in their genotype. In this example, all samples that had fluorescence lower than 400 fluorescence units were assigned as „negatives”. Negative amplification of blank controls, a positive signal in control DNA wells and clear peak of „positive” samples were basic quality criteria.

For the analysis of raw fluorescence a specialised software was created (available from http://www.mucosa.de/neu/index_bioinformatics.html). The software allows manual entering of cut-off values for the scoring of each PCR reaction. Negative, unknown and positive cut-offs were defined based on the signal intensity of the allele-specific positive and negative controls included in each plate as described previously. Where the signal intensity did not allow assignment of the result to either a positive or negative amplification, the sample was marked as unknown.

After all seven PCRs were performed and cut-offs for „positive” and „negative” samples were defined, final genotypes were assigned by software (see fig.8). In this example, the sample located at the position A1 has the genotype „3/5”. This result comes from positive amplifications in the amplification groups 3 and 5 and negative amplifications in groups 1,2,4,6 and 7.

Fig.8. Screenshot of the "DRB tool" software. The software displays the final composite genotypes for a 96-well plate. Discrepancies that occur during the verification of Mendelian inheritance, positions of positive and negative controls are highlighted automatically in red.

The screenshot shows the 'DRB Group Typing' software window. At the top, there is a 'Plate:' dropdown menu set to 'K16'. Below this are two tabs: 'Group Readings' and 'Genotypes', with 'Genotypes' being the active tab. A 'Make Genotypes' button is located above the data table. The table displays genotypes for rows A through H and columns 1 through 12. Some cells contain red text, indicating discrepancies or control positions.

row	1	2	3	4	5	6	7	8	9	10	11	12
A	35	35	13	13	3	3	3	23	3	23	3	12
B	5	34	14	3	13	3	35	35	14	13	24	25
C	2	13	2	13	25	23	3	23	35	2	12	4
D	5	2	1567	23	35	34	2	13	24	1567	3	3
E	3	15	2	34	3	35	3	3	23	23	35	3
F	3	12	34	3	35	13	3	23	3	3	13	13
G	23	3	1	12	23	1	15	3	3	47	3	1
H	2	34	23	35	3	1567	3	35	35	3	15	1567

Finally, the data was tested for inconsistencies (e.g. false positive amplifications creating the final „genotype” with three or more chromosomes) and Mendelian segregation (possible only in pedigrees). The software facilitated import of the data into the laboratory information system¹³⁰.

This method was applied to the entire sarcoidosis sample and the healthy unrelated individuals. A genotyping success rate of 94.11% was achieved. Genotyping failure was recorded in situations where signal intensities could not be clearly assigned either to the positive or negative groups. A percentage of 1.3 of the data had to be discarded from the family data due to violations of Mendelian inheritance within the families. This error rate compares favorably to other genotyping methodologies, for example Matise et al.¹³¹ reported recently a 1.6% error rate for SNP data. The analysis of genotype and allele frequencies in the *DRB1* allele-group genotyping data revealed no departures from Hardy-Weinberg Equilibrium.

2.7. SNP verification and mutation detection

The successful design of TaqMan™ assays for an association study requires detailed analysis of the DNA sequence in a set of individuals. This typically consists of a gene content analysis in the region and a search for annotated sequence variations in the genomic sequence data bases (in <http://www.celera.com/>; <http://genome.ucsc.edu/> and <http://www.ncbi.nlm.nih.gov/SNP/index.html>). Known SNPs may then have to be validated and/or patient DNA screened for novel polymorphisms.

2.7.1. SNP selection

SNP selection criteria depend on the aims of the study. The purpose of screening the chromosome 6 linkage region with biallelic markers was to identify the association region. 16.34Mb of linkage region were screened with 69 SNPs from the marker set developed for identification of disease-susceptibility genes located in the peri-HLA region¹¹⁷ (for marker set see table 8). The average density of the markers in the screening region was one per 240kb. The highest marker density was achieved in the linkage peak region (e.g. four SNPs were located in the *TNFA* promoter region), whereas more distant regions had lower marker density (e.g the maximal distance was between markers rs943461 and rs755658 located at 33.77 Mb and 35.55Mb respectively).

The 69 used assays for the I stage chromosome 6 screen were already optimised and typed on at least two 96 well plates with unrelated individuals from control sample for assay quality validation and Hardy-Weinberg equilibrium testing (Hardy-Weinberg equilibrium was re-estimated later in the entire sample of controls). Two candidate association regions were identified, one at *MICB-IkBL* (demonstrating only case-control association) and one at *BTNL2-DRA* (demonstrating consistent results in both TDT and case-controls analyses). These loci were analysed for gene content and sequence variation. A set of pre-available Assays -on-Demand and Assays -by-Design (both from Applied Biosystems Inc., Foster City, CA, USA) was checked for available markers covering the both regions and adding one flanking gene upstream and downstream: *MICA-TNFA* (rs1063635-rs1799964) and *AGER-DQAI* (rs1035798-rs707952). A high density map was established at the *AGER-DQAI* locus by selecting 48 markers covering 440kb (32.177Mb to 32.616Mb in NCBI release 32). Because of the complex patterns of haplotype structure and high sequence variation in the

HLA region, densely located biallelic markers were established in the genomic sequence between the *AGER* (rs2070600) and *HLA-DRA* (rs2395182). The average distance between these 47 markers in that region was 5.6kb. Distance to the last marker of the high-density LD map, *HLA-DQAI* (rs707952), varies according to individual HLA haplotype structure.

In addition to pre-available Assays-on-Demand and Assays-by-Design (Applied Biosystems Inc., Foster City, CA, USA), additional SNPs were established through *de novo* mutation detection and through verification of annotated sequence variations. A targeted SNP search was performed on the candidate *BTNL2* gene and the flanking intergenic sequences, where a higher density of markers was required. Non-synonymous SNPs found in the *BTNL2* coding sequence had a priority. In cases where only synonymous or non-coding polymorphisms were found, SNPs with a higher frequency and therefore informativity were selected for the further typing.

SNP markers were genotyped in the *CARD15* gene, a candidate gene for sarcoidosis association. Selection of these markers was based on the association findings in Crohn's disease. The *CARD15* sequence variations in the NOD domain, which is responsible for Blau syndrome, were investigated by sequencing 47 sarcoidosis DNA samples.

2.7.2. Primer design and PCR

The primers for mutation detection were designed for sequencing the exons and flanking intronic sequences in candidate genes, known SNP verification or searching for new SNPs in intergenic regions at loci of interest. Primers for exonic sequences were generally located in introns, covering the exon itself, splice site and 20-50bp of intronic sequence. Typically, amplicon length was 500-700bp. For longer exons, overlapping primer pairs were designed. Both chromosomes were always sequenced to avoid sequencing artefacts and verify the SNP on both the forward and the reverse strand. Primers were designed using the Primer Express 2.0 software (Applied Biosystems Inc., Foster City, CA, USA). If the program has failed to find optimal primers automatically, the oligonucleotide sequences were selected manually and tested for self-annealing loops and dimers.

PCR optimisation was carried out according to manufacturer's protocols (Qiagen GmbH; Hilden, Germany). Primers were diluted with DDW to a storage concentration of 100pmol/μl

and a working concentration of 20pmol/μl. The reaction solution contained the following reagents: 2.5μl Qiagen Buffer (10x), 0.5μl MgCl₂ (25mmol), 0.5μl dNTP (10mmol each), 0.25 forward primer (20pmol/μl), 0.25 reverse primer (20pmol/μl), 0.2μl Taq polymerase (Qiagen GmbH; Hilden, Germany) and DDW to a final volume of 25.0μl per reaction. 2.5ng dried or liquid human DNA was amplified. MgCl₂ was titrated for problematic amplifications. The reaction solution was set up for 13 reactions (of which 12 were used) and pipetted into the MicroAmp®single strips (Applied Biosystems, Darmstadt, Germany) or Microtiter 96-well plates (for testing up to eight primer pairs) (Costar Corning Incorporated; Cambridge, MA, USA) in the T1 Gradient (Whatman Biometra GmbH; Göttingen, Germany). An annealing temperature gradient was used to select the optimum annealing temperature for each primer pair. A two step touchdown program was employed with denaturation at 95°C for 10 min followed by 16 cycles of denaturation at 95°C for 30 sec, annealing at 66°C (the middle annealing temperature of the gradient; reduced by 0,5°C with every following cycle) for 30 sec, and extension at 72°C for 30 min. This was continued by 19 cycles of denaturation at 95°C for 30 sec, annealing at 58°C (the middle annealing temperature of the gradient) for 30 sec, extension at 72°C for 30 sec and final extension at 72°C for 10 min. The gradient covered annealing temperatures 5°C lower and higher than the given conditions. 5μl of PCR amplification product together with 5μl of 2times loading buffer (0.25% bromphenol blue, 0.25% Xylene Cyanol FF, 30% Glycerol in Water) were loaded into a 2% agarose gel (300ml TBE (Tris borate EDTA), 3μl ethidium-bromide). A 100bp ladder (Invitrogen, GmbH; Karlsruhe, Germany) was run to size the PCR products. PCR products were run at 150V for 50 min. The gel was analysed under UV light. A single clear band of expected size and only a low amount of primer-dimer indicated optimal PCR conditions. As necessary, optimisation was repeated with either a modified PCR mix (MgCl₂ concentration gradient ranging from 0.5μl to 3.0μl per reaction) or modified cycle conditions (changing annealing temperature).

2.7.3. Sequencing

For SNP identification, DNA from mutation detection patients (see chapter 2.3.3. for patient information) was amplified using the reaction conditions optimal for each primer-pair. The reaction was performed in the GeneAmp®PCR system 9700 (Applied Biosystems; Darmstadt, Germany) and visualized on a 2% agarose gel as described in the chapter 2.7.2. The criteria for the PCR quality were a single band of amplified product with a low amount of primer-dimer, a low number of failed reactions in the patient samples and no visible

amplification in the blank control well. Following PCR an enzymatic digestion step removed primer-dimers and free dNTPs from the samples. If the band on the gel was very intense, the amplified PCR product was diluted with the DDW (1:5). For weak bands, the original PCR product was used. The digestion reaction contained 0.30 μ l SAP (Shrimp Alkaline Phosphatase; 1U/ μ l), 0.15 μ l ExoI (Exonuclease I; 10U/ μ l) and 1.55 μ l DDW mixed with 8 μ l of PCR product (either concentrated or diluted) and was transferred to a new plate. The samples were incubated for 15min at 37°C followed by 15min at 72°C. Sequencing was performed using BigDye™ chemistry (BigDye Terminator Kit, Applied Biosystems, USA) according to the manufacturer's protocol. Next, in a separate plate, 2 μ l of the digested product was mixed with 1.0 μ l forward primer (1.6pmol/ μ l), 1.0 μ l of Big Dye Ready Reaction Mix and 6.0 μ l DDW. The same reaction was prepared with the reverse primer. The samples were incubated at 95°C for 5min followed by 25 cycles at 95°C for 1min, 50°C for 45sec and 60°C for 4 min. The thermal cycle was completed with a 5min extension at 60°C. After sequencing, samples were purified using Sephadex spin columns. The Sephadex powder (G50 Superfine; Amersham Biosciences; Freiburg, Germany) was measured with a 96 well format Multiscreen Column loader and applied to special filter plates (MAHVN 4550). 300 μ l of DDW were added into each well. The filter column plates were then incubated for ~3 hours at a room temperature followed by centrifugation at 910xg for 5min to eliminate excess water. The columns were then washed with 150 μ l of DDW and again centrifuged at 910xg for 5min. Finally, the sequencing product was mixed with 20 μ l of water and applied to the columns. The plate was centrifuged at 910xg for 5min the purified sequencing product collected in a MicroAmp® Optical 96-Well Reaction Plate (Applied Biosystems Inc., Foster City, CA, USA). Sequence detection was performed with an automated, high-throughput, 96-capillary fluorescence detection system ABI PRISM®3700DNA Analyzer (Applied Biosystems Inc., Foster City, CA, USA). The sequences of both strands were analysed and compared to the consensus sequence using the Sequencher program (Version 4.0.5. Gene Codes Corporation, Ann Arbor, MI, USA).

2.7.4. SNP verification and mutation detection in the association region and in sarcoidosis candidate genes

2.7.4.1. The BTNL2 locus on chromosome 6

Taking the consistency of the TDT, case-control results and significance levels into account, a ~15kb region at the 3-prime end of the *BTNL2* gene (from marker hCV2488476 to hCV2488463) yielded results two to three orders of magnitude more significant than the other regions. The *BTNL2* gene is located within the linkage region and within the associated region identified in the association mapping experiment. *BTNL2* therefore qualified as a candidate gene based on this data. Consequently, mutation detection of all exon sequences and the adjacent intron boundaries of *BTNL2* was performed in 47 sarcoidosis patients. Alignment of the revised cDNA sequence (see sequence information in the supplement 9.1.8.) with the genomic sequence (AL935032) was performed to determine the exact positions of the exons. Sequencher (version 4.0.5 created by Gene Codes Corporation, Ann Arbor, MI, USA) was employed for the analysis of all experimental sequences. Annotated sequence variations were compared to the experimental sequences and new SNPs were recorded. The coding sequence was located between 32.43Mb and 32.45Mb on chromosome 6p21 in the NCBI (release 32).

The *DRA* promoter, which in contrast to the coding region of this HLA molecule has not been hitherto investigated for polymorphisms, was sequenced in the same group of patients. To achieve a high map-density, markers in the genomic sequence between *BTNL2* and the *TSBP* were established either by direct sequencing or retrieved from the databases. The oligonucleotide primer sequences and reaction conditions used for amplification and sequencing of genomic DNA are listed in the supplement 9.1.6.

2.7.4.2. CARD15 on chromosome 16

Based on haplotype information from a genome-wide linkage analysis in 63 families⁵⁸, 39 sib pairs or trios were identified as sharing at least one parental chromosome 16. One patient representing each of these families, together with eight patients who represent seven complex families and do not share the parental chromosome 16, were selected for DNA sequence analysis of *CARD15* exon 4. In order to cover 1175 bp of exon 4, including the sites of

missense mutations responsible for susceptibility to Blau syndrome^{132, 133}, two overlapping primer pairs were used:

5'-TCCCTTCAGTTATGTCAGCGTC-3' / 5'-CAGAGAAGCCCTTGAGGTTGA-3'

and

5'-AAACCACTCTCTGTGCGGACTC-3' / 5'-AGAACGCGGCAAAGAAGCA-3'.

2.8. Internal data bases

An integrated internal data base system¹³⁰, with data entry and retrieval tools, was employed to store patient, genotype and assay information. All samples were entered into a patient data tables of the data base and received a unique identification number. Pedigrees contained all available information about probands and their relatives: phenotypic trait (affection status), date of birth and gender.

Plate layouts were also stored in the data base. Information about the sample positions on 96- and 384-well format plates was stored and plates were identified with unique numbers and labelled using a barcode system. Samples were assigned to certain positions in the plate layout.

Genotype data and assay information was also stored. After the assay was run, the resulting genotypes were entered into the data base and the data linked to the barcode on the plate. If families were genotyped, internal software checked for Mendelian inheritance errors. In cases where inconsistencies were found, genotypes for that entire pedigree at the assay in question were excluded from the data base. Primers and probes data were also stored in the database along with the assay reaction conditions. The data base tools also permitted the construction of user-defined marker maps, and marker sets for data export and further analysis.

A multiallelic-genotype information tool was developed for the DRB1 allele-group genotyping experiment. This tool was developed for definition of allele-group positive and negative amplification in samples, genotype construction and checking of genotyping inconsistencies. Information was linked to the main genotype data base described above thus enabling the concurrent export of multiallelic and diallelic genotypes.

All regions of the data base were relationally linked to each other thus integrating all available sample information. Additional software programs were used for the controlling of robots used in processing DNA and assay preparation. The genotype data base was facilitated the production of export files with sample genotype information for detailed statistical analysis programs running under UNIX.

2.9. Statistical analysis

Prior to detailed genetic analyses, each marker was tested for Hardy-Weinberg equilibrium¹³⁴,¹³⁵ in the control population using a χ^2 test.

Many analyses utilized utility scripts implemented in the GENEDIGGER software environment (www.genedigger.de) for genetic analysis using the principles outlined below.

Single point and haplotype transmission disequilibrium tests¹³⁶⁻¹³⁸ were performed for association analysis using TRANSMIT¹³⁹ and GENEHUNTER 2.1¹⁴⁰. In cases of families with multiple affected individuals, analyses were performed by random selection of single trios from those families (as implemented by both programs). Haplotype frequency estimates of singletons were obtained using an implementation of the expectation maximisation (EM) algorithm (HAPMAX, www.uwcm.ac.uk/uwcm/mg/download)¹⁴¹. Significance testing of haplotype frequency differences was performed with HAPMAX, making use of the fact that twice the log-likelihood ratio between two nested data models approximately follows a χ^2 distribution with k degrees of freedom, where k is the difference in parameter number between the two models. Significance assessment of associations with or between single locus genotypes was performed with χ^2 or Fisher's exact test for 2 x 3 contingency tables (<http://home.clara.net/sisa/fishrhlp.htm>). Allele and genotype frequency differences between cases and controls were calculated using Pearsons χ^2 (<http://home.clara.net/sisa/two2hlp.htm>).

In order to assess the relative importance of the *MICB* and *BTNL2* leads, logistic regression analysis was used. Genotypic regression analysis was performed with SPSS analysis (<http://www.spss.com/>). Individual SNP genotypes were coded as categorical variables and the affection status was used as a binary outcome measure. Forward likelihood ratio inclusion was used in the program.

Standardized coefficients r^2 and D' ^{8, 26} were computed for pair-wise LD estimation between markers. To this end, haplotype frequencies were estimated using the EM implementation in HAPMAX, excluding markers that frequency was below 1%¹⁴¹. Breslow-Day test (http://www.cochrane.dk/cochrane/handbook/hbookBreslowDay_test.htm) for homogeneity of odds ratios of genotypes was used to stratify the *BTNL2* rs2076530 effect from the *DRBI* gene allele-groups.

3. Results

3.1. Association study in the peri-MHC

3.1.1. Identification of the association lead on chromosome 6

For the association study of sarcoidosis susceptibility genes on chromosome 6 we defined the 27.8-44.1 Mb region on the basis of linkage as identified in a whole genome scan (fig. 9A according to published information by Schürmann et.al.⁵⁸). The „exploratory” sample of sarcoidosis patients was genotyped for 69 SNP markers covering the linkage region that was defined by a decrease in LOD score by 1.5 units or less, representing an approximate 99% confidence region for linkage in the original genome screen⁵⁸ (see marker details in supplement 9.1.1.). In order to maximise the informativity of the data, three-locus SNP haplotypes were analysed in the TDT (fig. 9B) and case-control (fig. 9C) designs. A P-value less than 0.001 was selected as a cut-off for our association mapping study. According to this criterion, only the *BTNL2-HLA-DRA* locus (marker rs2076523 to rs7192) demonstrated a consistent association signal in both study designs. A second putative peak was detected in the *MICB-IkBL* sub-region. This region was significant only in the case-control analysis. Nevertheless, the peak was further investigated by typing additional markers (rs1063635, rs3134900, rs3130062, rs1041981 and rs1799964) in the „extended” sample of sarcoidosis patients (see table 3 for patient information). In the same sample set, markers rs2076523, hCV2455668, hCV2455646, rs7192 were genotyped to verify the association signal in the *BTNL2-HLA-DRA* region. Genotype logistic regression analysis with likelihood ratio-based forward inclusion at the $p < 0.05$ level resulted in a model that included only two markers from the *BTNL2-DRA* region (rs2076523, rs7192), but none of the *MICB* markers. Furthermore, long range LD with D' values up to 0.94 was observed between markers of the two loci (see fig.10). Thus, further association study was pointed only to the *BTNL2-HLA-DRA* locus.

Fig.9. Graphical representation of the stage I SNP screen on chromosome 6p21. Panel A illustrates the NPL-score curve from the original genome scan⁵⁸. Approximately 99% of the confidence interval for linkage is marked in grey. Panel B: results of a three-locus haplotype TDT, using TRANSMIT¹³⁹. Panel C: results of a three-locus haplotype case-control analysis, using HAPMAX¹⁴¹. In panels B and C, the black dots represent markers; the P-value is plotted against the physical map position (NCBI coordinates; release 32) of the most telomeric marker of each haplotype. Using a cut-off of p-value 0.001, two potential lead regions (*MICB* and *BTNL2* regions) were identified (marked by arrows).

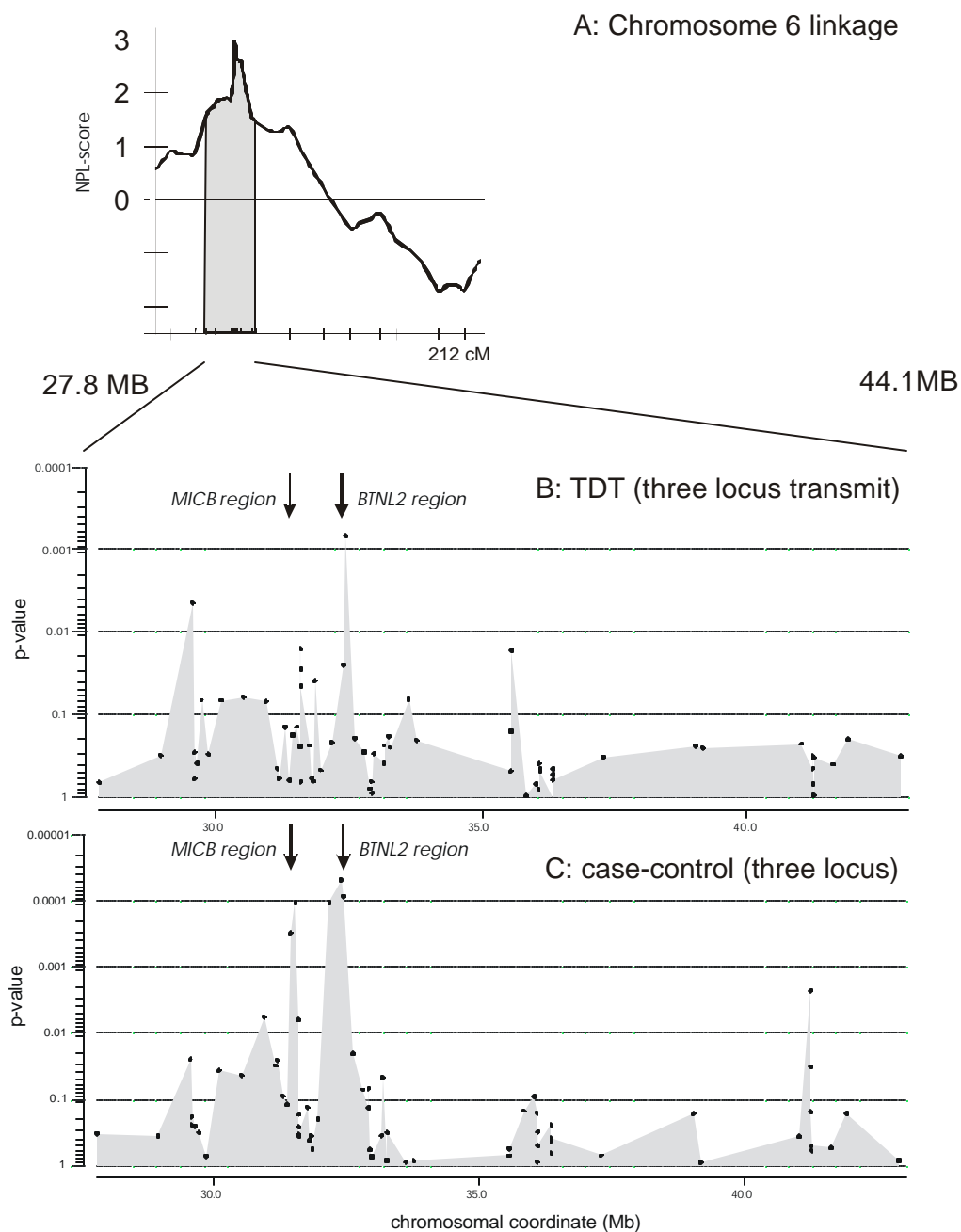


Table 8.Part 1: Analysis of 69SNP marker covering chromosome 6 linkage region typed in the „exploratory” sample. Three-locus haplotype analysis: haplotype TDT, using TRANSMIT¹³⁹. and three-locus haplotype case-control analysis, using HAPMAX¹⁴¹. Meaning of the headers- Marker: name of the first marker in the analyzed three-marker haplotype; Position: map position of this marker in NCBI release 32; Gene: gene name is provided, if the marker was assigned to the gene; GLOBAL_P: global P-value (using tabulated chi-squared values) of the TRANSMIT analysis, (represented graphically in the B panel of fig.9) LR_P: likelihood ratio P-value in HAPMAX analysis (represented graphically in the panel C of fig.9). Using a cut-off of P-value 0.001, two potential lead regions (*MICB* and *BTNL2* regions) were identified. Three-locus analysis results are represented at the first marker, thus no separate information can be provided for the last two SNPs.

Nr	Marker	Gene	Position MB	GLOBAL_P	LR_P
1	rs126007		27,80	0.66272	0.33685
2	rs929042	<i>RFP</i>	28,95	0.31158	0.35774
3	rs53161		29,56	0.00439	0.02515
4	rs404240	<i>UBD</i>	29,59	0.60590	0.24800
5	rs362536	<i>UBD</i>	29,59	0.28863	0.18665
6	rs1805057	<i>GABBR1</i>	29,64	0.38716	0.25904
7	rs3130253	<i>MOG</i>	29,73	0.06654	0.32363
8	rs1569315	<i>HLA-G</i>	29,85	0.30168	0.73116
9	rs8347	<i>HCG-V</i>	30,10	0.06731	0.03672
10	rs1264457	<i>HLA-E</i>	30,52	0.06068	0.04346
11	rs2074510	<i>GTF2H4</i>	30,94	0.07043	0.00569
12	rs3095318	<i>CDSN</i>	31,15	0.44813	0.03073
13	rs2073721	<i>SC1</i>	31,19	0.59125	0.02606
14	rs1050393	<i>HLA-C</i>	31,30	0.14277	0.08974
15	rs1051488	<i>HLA-B</i>	31,38	0.62729	0.11985
16	rs1063635	<i>MICA</i>	31,44	0.17855	0.00031
17	rs3134900	<i>MICB</i>	31,53	0.14103	0.00011
18	rs3130062	<i>IkBL</i>	31,59	0.24477	0.00631
19	rs1799964	<i>TNFA</i>	31,60	0.65624	0.17068
20	rs1799742	<i>TNFA</i>	31,60	0.02797	0.36793
21	rs1800629	<i>TNFA</i>	31,60	0.01595	0.26435
22	rs361525	<i>TNFA</i>	31,60	0.04533	0.34753

Table 8.Part 2: Analysis of 69SNP marker covering chromosome 6 linkage region typed in the „exploratory” sample.

Nr	Marker	Gene	Position MB	GLOBAL_P	LR_P
23	rs707916	<i>DDAH2</i>	31,76	0.23426	0.13348
24	rs707926	<i>VAR2</i>	31,81	0.59715	0.42364
25	rs2075800	<i>HSP70-HOM</i>	31,84	0.64525	0.36278
26	rs539689	<i>HSPA1B</i>	31,86	0.03912	0.57590
27	AF019413_C89532G	<i>C2</i>	31,96	0.46907	0.20283
28	rs1035798	<i>AGER</i>	32,18	0.21567	0.00011
29	rs2076523	<i>BTNL2</i>	32,40	0.02512	0.00005
30	rs7192	<i>HLA-DRA</i>	32,44	0.00068	0.00008
31	rs707952	<i>HLA-DQA1</i>	32,62	0.19542	0.02065
32	X66401_G48629A	<i>TAP2</i>	32,80	0.28180	0.07202
33	rs1042337	<i>HLA-DMB</i>	32,91	0.78721	0.13677
34	rs1063478	<i>HLA-DMA</i>	32,92	0.64879	0.06935
35	rs3918143	<i>RING3</i>	32,95	0.89879	0.58044
36	rs2581	<i>HLA-DNA</i>	32,98	0.29726	0.74512
37	rs926424	<i>RXRB</i>	33,16	0.39056	0.35662
38	rs1547387	<i>HKE4</i>	33,17	0.23682	0.04729
39	rs466384	<i>BING4</i>	33,26	0.18172	0.85135
40	rs2071888	<i>TAPBP</i>	33,27	0.24968	0.32265
41	rs999942	<i>ITPR3</i>	33,63	0.06519	0.89534
42	rs943461	<i>MLN</i>	33,77	0.20307	0.86884
43	rs755658	<i>FKBP5</i>	35,55	0.48398	0.72043
44	rs992105	<i>FKBP5</i>	35,56	0.15834	0.72385
45	rs1320991	<i>FKBP5</i>	35,56	0.01655	0.57041
46	rs1326752	<i>SRPK1</i>	35,83	0.97381	0.14943
47	Z95152_G57723A		36,03	0.69303	0.09150
48	Z95152_T58784G		36,08	0.79999	0.16206
49	Z95152_T106861C		36,08	0.39651	0.90485
50	rs763096	<i>MAPK13</i>	36,10	0.44718	0.31687
51	rs2007683	<i>MAPK13</i>	36,10	0.48579	0.51913

Table 8.Part 3: Analysis of 69SNP marker covering chromosome 6 linkage region typed in the „exploratory” sample.

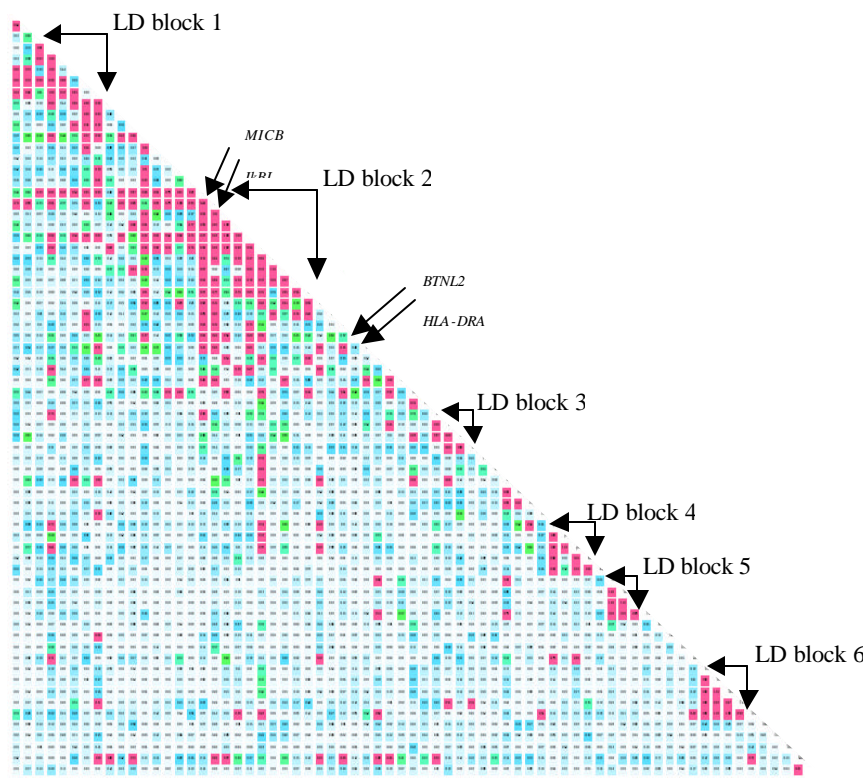
Nr	Marker	Gene	Position MB	GLOBAL_P	LR_P
52	rs881648	<i>TEL2</i>	36,35	0.39159	0.24957
53	rs881647	<i>TEL2</i>	36,35	0.52845	0.43205
54	rs916287	<i>TEL2</i>	36,35	0.45249	0.66488
55	rs1885205	<i>TEL2</i>	36,35	0.60848	0.39376
56	rs2055272	<i>FLJ20337</i>	37,29	0.32889	0.71445
57	rs1004279	<i>GLP1R</i>	39,03	0.23950	0.16773
58	rs1541816	<i>KCNK5</i>	39,16	0.25399	0.91441
59	rs1014737	<i>APOBEC2</i>	41,02	0.22457	0.36773
60	rs2234246	<i>TREM1</i>	41,25	0.45006	0.00228
61	rs1817538	<i>TREM1</i>	41,25	0.69467	0.03263
62	rs1351835	<i>TREM1</i>	41,25	0.32326	0.15618
63	rs2234237	<i>TREM1</i>	41,25	0.95560	0.60305
64	AL391903_C88694T	<i>TREM1</i>	41,26	0.33000	0.52120
65	rs742735	<i>MDF1</i>	41,62	0.39961	0.55194
66	rs1410492	<i>CCND3</i>	41,91	0.19828	0.16327
67	rs1160867	<i>TNRC5</i>	42,90	0.32091	0.85330
68	rs9395	<i>SRF</i>	43,15		
69	rs1418488	<i>CAPN11</i>	44,14		

3.1.2. Analysis of LD in the screened region

Due to linkage disequilibrium between genes, a distortion of the true location of a susceptibility gene is likely. Therefore, pair-wise LD analysis in the control population was performed on the set of markers employed for the screening experiment. Generally, the lower is the frequency of an SNP, the further extended LD around such alleles is expected, because such alleles are generally young and there has been less historical opportunity for recombination to break down ancestral haplotypes¹⁴². Markers with a higher frequency than 1% were included into analysis of LD structure in this study. D' values were calculated for all pair-wise inferred two-locus haplotypes for the 69 SNPs covering a 16,34 Mb region. The LD

in this region was found to vary greatly, with strongly correlated regions of sequence forming so called LD blocks interspersed with regions of low LD. D' values greater than 0.5 were considered as evidence of linkage disequilibrium between markers¹³ (see fig.10). In the analysed fragment, six such blocks were discovered. In general, the telomeric end of the region presented higher LD values which are represented by „LD block 1” (29.55Mb-29.84Mb) and „LD block 2”(31.14Mb-31.85Mb). The strongest LD was observed between markers in the second region, which is bordered by rs3095318 and rs539689. This region includes the *MICB* and *IkBL* genes that exhibit high LD values over long physical distance reaching the *BTNL2-HLA-DRA* region. Interaction between the two putative association regions was of a special interest because both regions demonstrated significant results in the case-control analysis. The size of LD blocks 3 to 6 was notably smaller, ranging from 253bp (LD block 5) to 91kb (at LD block 3). These blocks represented densely located markers. The LD analysis results in our sample set in the peri-MHC region agree with previous studies that have asserted that apart from the classical HLA loci¹⁹, the LD patterns in this region are no different from those found in other studies in the rest of the genome^{13, 14, 20, 21}.

Fig10. Linkage disequilibrium between 69SNP markers typed for association screening on chromosome 6. 16,34Mb sequence of 517 unrelated individuals of German origin were tested for LD between markers. The 69 SNPs are plotted against each other (from left to right and from top to bottom). The colour squares code D' values. LD values from 0.00 to 1.00 are colour-coded as following: light blue- blue (0.0-0.34), green (0.35-0.49) and red (≥ 0.5). „Blocks” formed by D' values ≥ 0.5 were designated as „LD blocks1-6”, formed by markers 3-8, 12-26, 37-39, 47-50, 52-54 and 60-63 respectively.



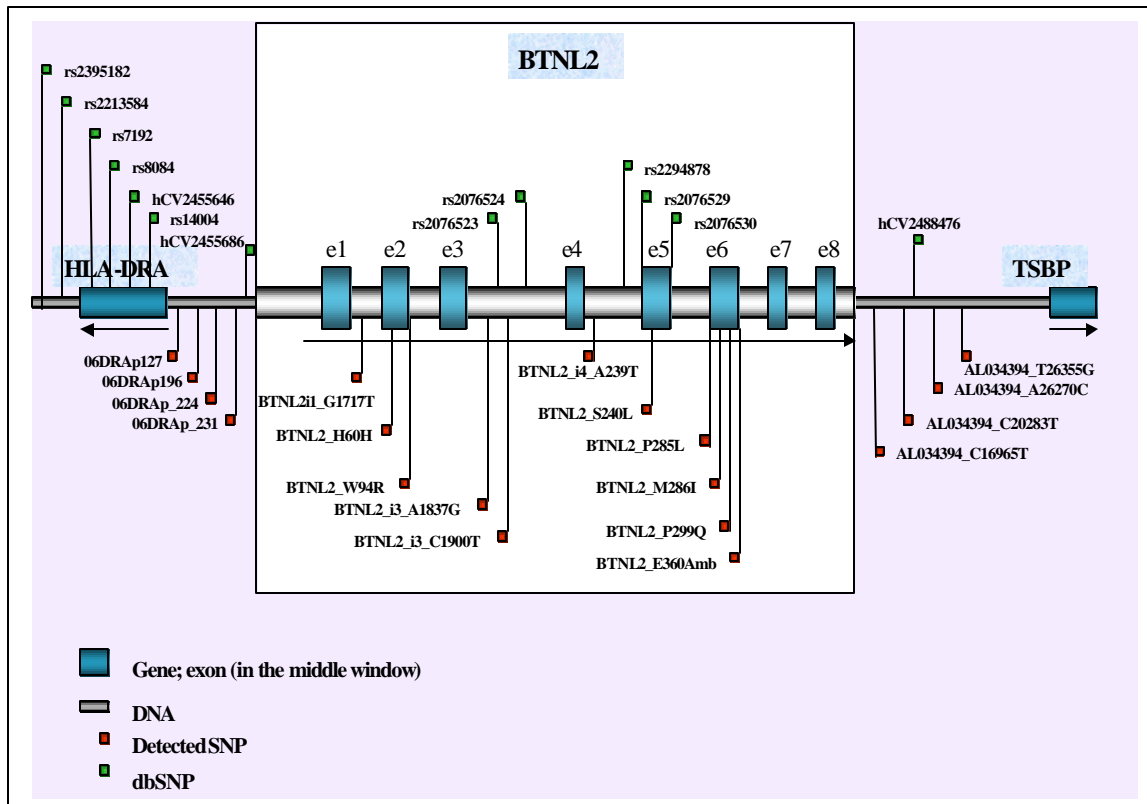
3.1.3. High density SNP mapping in the association region

Further investigation of the *BTNL2*-*HLA-DRA* region comprised genotyping 48 additional markers (NCBI release 32, map coordinates 32,177 – 32,616 Mb) in the „fine mapping” sample. This sample consisted of the combined „exploratory” and „extension” samples (see table 3. for sample information).

To establish a dense SNP map in the locus and to identify the disease-associated marker, 47 sarcoidosis patients were sequenced for mutation detection screening (see fig.11). The

sequencing experiment covered all *BTNL2* gene exons and adjacent intronic sequences. Seven non-synonymous polymorphisms were identified: BTNL2_W94R (tryptophan changed into arginine at amino acid position 94 of the revised BTNL2 transcript see sequences in supplement 9.1.7.); BTNL2_S240L; rs2076530; BTNL2_P285L; BTNL2_M286I; BTNL2_P299Q; BTNL2_E360Amb. Two synonymous SNPs were identified: BTNL2_H60H; rs2076529. Seven intronic sequence variations were identified: BTNL2_i1_G1717T (G nucleotide is changed into T at position 1717 of intron 1 of the BTNL2 gene); BTNL2_i3_A1837G; rs2076523; BTNL2_i3_C1900T; rs2076524; BTNL2_i4_A239T; rs2294878. These SNPs were selected for high density SNP mapping (see marker details in supplement 9.1.1.). Further SNPs were established in the intergenic sequence between the *BTNL2* and *TSBP* genes: AL034394_C16965T (C changed to T at nucleotide position 16965 of sequence AL034394 from the BAC clone RP5-107715); AL034394_C20283T; hCV248876; AL034394_A26270C and AL034394_T26355G (see fig.11). The remaining markers were located in the *AGER*, *NOTCH4*, *TSBP* and *HLA-DRA* genes and were obtained from Assays-on-Demand and Assays-by-Design (Applied Biosystems Inc., Foster City, CA, USA) or sequencing experiment.

Fig.11. Fragment of the high density SNP map established in the association region.
 Upper panel: SNP markers (marked in green) identified through ABI and dbSNP databases.
 Lower panel: SNP markers established through mutation detection experiments. Central panel: SNP marker locations in the revised *BTNL2* gene; e: exon.



Additional assays were obtained either from the ABI „Assay on Demand” database (www.store.appliedbiosystems.com), or from the dbSNP web page (www.ncbi.nlm.nih.gov). Single point TDT and case-control analyses were performed to evaluate the contribution of individual markers to the significant signal detected at the locus. Taking the consistency of results from both the family- and population-based analyses into account, markers at the 3’ end of the *BTNL2* region (rs2294878 to hCV2488476) demonstrated the strongest and most robust association (see fig.12.). Average rate of the P-values for these markers were two to three orders of magnitude smaller than for other significant markers. The human BTN gene cluster consists of seven genes (*BTN1A1*, *BTN3A3*, *BTN2A1*, *BTN3A1*, *BTN2A3*, *BTN3A2* and *BTN2A2*) and is located on chromosome 6p22^{80, 143} approximately 6 Mb telomeric of *BTNL2*. For completeness, six markers (rs2073528, rs2072803, hCV2474907, hCV2474883, hCV9804528 and rs3736781) covering ~130kb of the BTN cluster were also genotyped in the combined “exploratory” and “extended” samples of the present study. There was no evidence for disease association (data not shown).

Fig.12.High density mapping in the association region. SNP markers comprising 440 kb region of chromosome 6p21 were typed in the combined „exploratory” and „extension” samples. Table: results of disease-association tests. f_{controls} : allele frequency in controls; TDT: P-values from a single point TDT calculated by GENEHUNTER 2.1, Case-control: single point P-values from a genotype-based χ^2 test of case-control data. Colour intensity-code was used to mark decreasing P-values. Right panel: revised intron-exon structure of the *BTNL2* gene. The most disease-associated *BTNL2* markers P299Q, M286I, P285L (located in exon 6) and rs2076530 (located in exon 5) are indicated with arrows.

marker	gene	f_{controls}	TDT	Case_control
rs2070600	<i>AGER</i>	0.03	0.5637	0.547556261
rs1800624	<i>AGER</i>	0.70	0.8932	0.888288117
rs2071287	<i>NOTCH4</i>	0.43	0.4857	0.471723107
rs384247	<i>NOTCH4</i>	0.13	0.7797	0.615679814
rs915894	<i>NOTCH4</i>	0.38	0.7963	0.050395291
rs436845		0.75	0.8282	0.170348756
rs560505	<i>TSBP</i>	0.57	0.1203	0.001818197
rs485774	<i>TSBP</i>	0.56	0.0872	0.001808805
rs552339	<i>TSBP</i>	0.44	0.1351	0.004321214
rs761187	<i>TSBP</i>	0.56	0.1482	0.003243398
rs2073044	<i>TSBP</i>	0.75	0.0107	0.000155551
rs2050190	<i>TSBP</i>	0.67	0.0993	0.019714402
rs2073045	<i>TSBP</i>	0.36	0.9468	0.038608552
AL034394_T26355G		0.93	0.3770	0.513599058
AL034394_A26270C		0.26	0.1509	0.000115550
AL034394_C20283T		0.03	0.7237	0.355508299
AL034394_C16965T		0.15	0.0066	0.001426440
hCV2488476		0.41	0.0001	0.000000021
AL034394_C12339T		0.15	0.0025	0.001453209
AL034394_T10552C		0.86	0.0027	0.002117374
BTNL2_E360Amb	<i>BTNL2</i>	0.99	0.4142	0.999962285
BTNL2_P299Q	<i>BTNL2</i>	0.15	0.0013	0.000565354
BTNL2_M286I	<i>BTNL2</i>	0.85	0.0008	0.001159376
BTNL2_P285L	<i>BTNL2</i>	0.85	0.0001	0.000186665
rs2076530	<i>BTNL2</i>	0.58	0.0009	0.000000103
BTNL2_S240L	<i>BTNL2</i>	0.98	0.2008	0.532618389
rs2076529	<i>BTNL2</i>	0.59	0.0001	0.000000044
rs2294878	<i>BTNL2</i>	0.56	0.0012	0.000000700
BTNL2_i4_A239T	<i>BTNL2</i>	0.97	1.0000	0.900004549
rs2076524	<i>BTNL2</i>	0.74	0.0588	0.000076997
BTNL2_i3_C1900T	<i>BTNL2</i>	0.92	0.5876	0.727167218
rs2076523	<i>BTNL2</i>	0.35	0.1385	0.000907369
BTNL2_i3_A1837G	<i>BTNL2</i>	0.08	0.3173	0.843149836
BTLIIW94R	<i>BTNL2</i>	0.08	0.9068	0.908877597
BTLIIH60H	<i>BTNL2</i>	0.92	0.7255	0.829924141
BTNL2_i1_G1717T	<i>BTNL2</i>	0.98	0.8084	0.993204521
hCV2455668		0.42	0.2359	0.041676362
DRAp_231	<i>HLA-DRA</i>	0.30	1.0000	0.451078993
DRAp_224	<i>HLA-DRA</i>	0.30	0.7400	0.492950227
DRAp_196	<i>HLA-DRA</i>	0.31	0.8435	0.490587034
DRAp_127	<i>HLA-DRA</i>	0.98	0.4927	0.984228521
rs14004	<i>HLA-DRA</i>	0.42	0.1761	0.058413905
hCV2455646	<i>HLA-DRA</i>	0.39	0.7622	0.937496596
rs8084	<i>HLA-DRA</i>	0.44	0.0083	0.000183610
rs7192	<i>HLA-DRA</i>	0.58	0.0014	0.000094646
rs2213584	<i>HLA-DRA</i>	0.59	0.0010	0.000176354
rs2395182	<i>HLA-DRA</i>	0.79	0.2626	0.000847780
rs707952	<i>HLA-DQA1</i>	0.28	0.0115	0.025177725

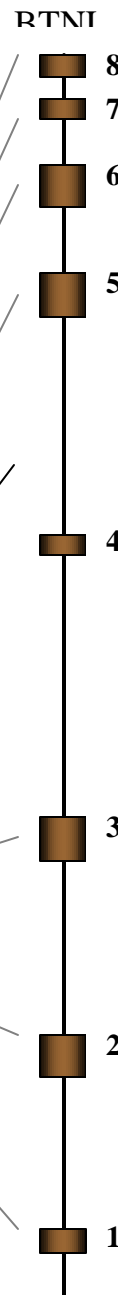
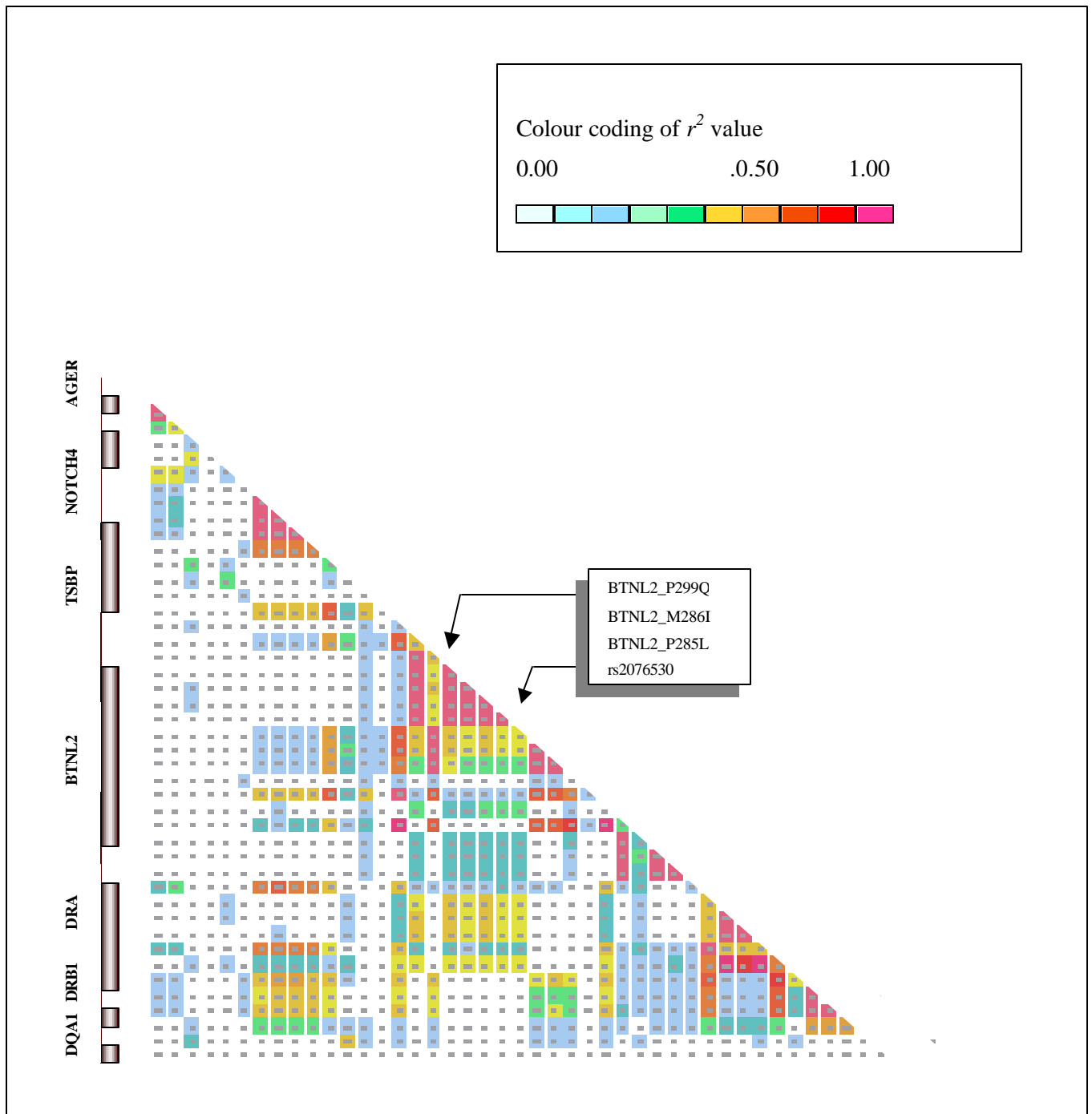


Fig.13. The matrix of pair-wise r^2 values estimated at the locus *AGER-HLA-DQA1* (from *rs1035798* through *rs707952*, as used in the stage II chromosome 6 screening). Left panel: gene map of the analysed 440 kb region on human chromosome 6p21. Equidistance plot of pair-wise LD (r^2) between all markers in the analysed region. 69 SNPs are plotted against each other (from left to right and from top to bottom). The left panel presents the physical map of the analysed region. The colour squares in the plot code r^2 values. r^2 values that range from 0.00 till 1.0 (no LD and complete LD respectively) are colour-coded according to LD strength (indicated in r^2 value legend). The most disease-associated marker positions in the LD plot indicated by arrows.



To investigate the association signal in more detail, the haplotype structure of the high density mapping region was analysed. According to our experience, D' values are of little utility between markers in close proximity, therefore a r^2 value of LD measure were investigated (see fig.13). The four potentially functional and with disease mostly associated markers rs2076530, P285L, M286I, P299Q had r^2 values ranging from 0.88 till 0.99 indicating high LD between markers.

From the 15kb segment, four SNPs (rs2076530, P285L, M286I, P299Q) in the *BTNL2* gene were the most strongly associated with sarcoidosis. Combined in a four marker TDT haplotype analysis they yielded a P-value of 1.9×10^{-7} , and a P-value of 6.2×10^{-7} in the equivalent case-control analysis. Most of the association was due to rs2076530, leaving $P=0.03$ for the inclusion of the other three markers in a case-control haplotype regression model. The strong association of rs2076530 with sarcoidosis was verified by typing a „replication” sample containing 452 independent cases (see table 9.). Both TDT and case-control analyses demonstrated increased significance in the „full” sarcoidosis patient sample, which represents the entire sample of 947 affected individuals (single affected individual per family). Transmission disequilibrium tests and case-control tests returned $P_{\text{tdt}} = 3.0 \times 10^{-5}$ and $P_{\text{case-control}} = 6.7 \times 10^{-8}$ values respectively.

Table 9. Significance of disease-associated *BTNL2* markers in the replication sample. The markers were analysed in the „SNP fine mapping” and „replication “ samples separately and combined. TDT p-value: single point TDTs were calculated by GENEHUNTER 2.1. Case control genotype p-value: single point genotype-based χ^2 test in the population study design. Homozygotes: odds of ratio between wild type homozygote genotype „2/2” and mutated type homozygote „1/1”; heterozygotes: odds of ratio between wild type homozygote „2/2” and mutated type heterozygote „1/2”. Marker rs2076530 demonstrated consistent results in both samples and had the highest odds ratio in homozygotes for allele „A”. Colour intensity indicated the decreasing P values.

Marker	Sample						Genotype odds ratio	
	SNP fine mapping		Replication		Full		Homozygotes	Heterozygotes
	TDT p-value	Case control genotype p value	TDT p-value	Case control genotype p value	TDT p-value	Case control genotype p value		
BTNL2_P299Q	0.00175	0.00056535	0.1118	0.644517	0.000631	0.024066193	0.61	0.71
BTNL2_M286I	0.00175	0.00115938	0.2752	0.429132	0.001963	0.008748047	1.65	1.11
BTNL2_P285L	0.00085	0.00018667	0.1129	0.269000	0.000322	0.006999268	1.86	1.26
rs2076530	0.00052	0.00000010	0.0019	0.000059	0.000003	0.000000067	2.59	1.53

The underlying allele transmission rates in the transmission disequilibrium tests and the allele and genotype frequencies in the case-control analyses of the marker rs2076530 are listed in table 10. The disease susceptibility allele, allele “1” (corresponding to the nucleotide A), was over-transmitted in families. 256 alleles were transmitted from heterozygous parents to affected children against 161 un-transmitted (a ratio of 1.59). Case-control statistics agreed with the TDT results with the following distribution of frequencies for genotypes AA, AG and GG in controls and cases respectively: 0.32; 0.53; 0.16 and 0.46; 0.44; 0.09.

Table 10. Association of the rs2076530 with susceptibility to sarcoidosis. Table A: transmission disequilibrium test. Trans: number of transmitted alleles; untrans: number of un-transmitted alleles. Table B: case-control analysis test. Frequencies of alleles and genotype frequencies are presented for various samples (see table 3 for detailed sample information). Allele1 denotes nucleotide A and allele 2 denotes nucleotide G of the rs2076530 marker.

A.

Allele	Sample					
	SNP fine mapping		Replication		Full	
	Trans	Untrans	Trans	Untrans	Trans	Untrans
1	158	102	98	59	256	161
2	102	158	59	98	161	256

B.

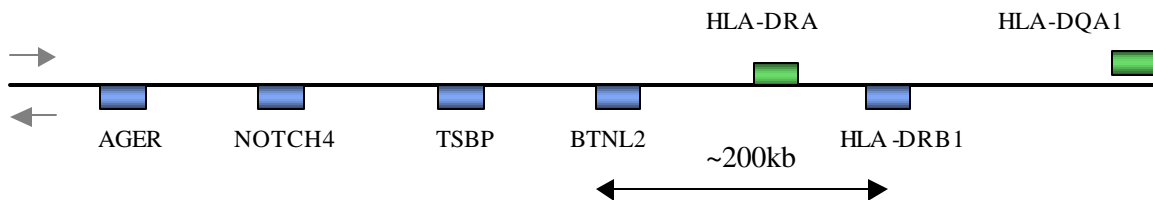
Alleles	Sample			
	Controls	SNP fine mapping	Replication	Full
1	0.58	0.70	0.68	0.69
2	0.42	0.30	0.32	0.31
Genotypes				
11	0.32	0.50	0.44	0.46
12	0.53	0.40	0.47	0.44
22	0.16	0.10	0.09	0.09

3.1.4. Separation of the genetic effect of *BTNL2* and the *HLA-DRB1* genes

The marker rs2076530 demonstrated high significance in family-based and population-based study designs. Thus, *BTNL2* could be regarded as a sarcoidosis susceptibility gene. However,

it is located in the vicinity of the *HLA-DRB1* gene (~200kb) (see fig.14), which has been previously implicated in the aetiology of the disease^{72, 74}.

Fig.14. Gene content in the fragment of the chromosome 6p21. The *BTNL2* gene maps ~200kb (indicated by black arrow) telomeric from *HLA-DRB1* gene. Grey arrows indicate gene direction.



To investigate the association signal and possible effect of linkage disequilibrium in the association region, the *HLA-DRB1* gene was genotyped. To differentiate the *BTNL2* genetic effect from the HLA marker, a low-resolution typing approach for non-overlapping *HLA-DRB1* allele groups was applied (see chapter 2.6.2. for experiment design details). Allele group genotyping results in the full sarcoidosis sample plus controls confirmed that a group of alleles confers an increased risk of sarcoidosis. Amplification group 3, comprising alleles *DRB1**03, *08, *11(not *1122/30), *12, *13, *14, (not*1410/39), demonstrated a significant signal (see table 11). To evaluate the relative genetic effect of both markers, a stratification analysis of the respective *HLA-DRB1* and *BTNL2* genotypes was employed. However, a stratified analysis of the respective *DRB1* and *BTNL2* genotypes (see table 12) was insignificant throughout (Breslow-Day test for homogeneity of odds ratios: $\chi^2=0.12$, 1 d.f., $p>0.5$ for both stratifications). Therefore, *DRB1* amplification group 3 and rs2076530 allele A represent independent risk factors for sarcoidosis.

Table 11: Association between *BTNL2* and *DRB1* risk alleles and sarcoidosis (based on 947 cases and 517 controls). f_{case} , f_{control} : allele frequency in cases and controls. P (global): error probability of a global χ^2 test for allelic association. tdt (obs./exp.) ratio between observed and expected number of transmissions; P (allele): allele group-specific TDT error probability, as reported by the TRANSMIT program¹³⁹.

Locus	allele	f_{case}	f_{control}	P (global)	tdt (obs./exp.)	p (allele)
<i>BTNL2</i> (rs2076360)	A	0.69	0.58		1303/1239	$3.1 \cdot 10^{-7}$
	G	0.31	0.42	$3.3 \cdot 10^{-9}$	605/669	$3.1 \cdot 10^{-7}$
<i>DRB1</i> (allele groups)	1	0.067	0.112		124/155	$1.1 \cdot 10^{-5}$
	2	0.193	0.169		344/330	n.s.
	3	0.551	0.461		1010/947	$7.4 \cdot 10^{-7}$
	4	0.100	0.146		186/217	0.0001
	5	0.082	0.103		145/156	n.s.
	6	0.003	0.005		10/4	n.s.
	7	0.004	0.004	$7.1 \cdot 10^{-8}$	7/3	n.s.

Table 12. Differentiation between *BTNL2* and *DRB1* effects. „x” denotes *DRB1* alleles other than allele group 3.

rs2076530	<i>DRB1</i>	case	control
any	3-3	282	125
any	3-x	481	227
any	x-x	184	165
AA	any	445	170
AG	any	414	261
GG	any	88	86
AA	3-3,3-x	388	137
AA	x-x	57	33
AG,GG	3-3,3-x	375	215
AG,GG	x-x	127	132

3.1.6. Functional exploration of *BTNL2* mutations

The association mapping experiments indicated four potentially functional SNPs in the *BTNL2* gene that are associated with sarcoidosis. Three SNPs are located in exon 6 and cause a proline change to leucine at amino acid position 285, methionine to isoleucine at 286 and proline to glutamine at 299 (denoted P285L, M286I and P299Q respectively). According to the original gene model, the A to G mutation at rs2076530 causes a serine to be replaced by glycine¹⁴⁴. However, in the revised gene model, the SNP is located at position -1 of the splice donor site, where guanine is conserved in 77% of donor splicing sites^{145, 146}. Thus, the substitution of this conserved nucleotide could influence exon splicing and alter the gene transcript.

To investigate the hypothesis of alternative splicing, DNA/cDNA pairs from lymphoblastoid cell lines and peripheral blood samples were compared by rtPCR (see method description in appendix 9.2.2.). Indeed, a four base deletion in the cDNA was observed in cases where the genomic DNA contained allele A at the -1 splice donor site (see fig.16). The loss of four bases in the *BTNL2* transcript causes a frameshift and generates a premature stop codon in exon 6. In splice variants with frameshift and premature stop codon, the 118 amino acids at the C-terminus are then replaced by 5 different residues.

Fig.16. Alternative splicing of the *BTNL2* transcript at exon 5 (experiment performed in collaboration with Dr.K.Huse and Dr.P.Rosenstiel). Alternative splicing is introduced through an A to G transition at rs2076530. Top panel: reference genomic sequence around the mutation site. Central panel: consensus splicing in the presence of the G allele. Bottom panel: alternative splicing in the presence of the allele A. The GT of the active donor is underlined in both panels.

	rs2076530
	↓
Reference sequence	GATCTGAAGGTGGTAR GT AAGAATTCTAGAT
cDNA	gatctgaaggtggta g ...
Genomic sequence	GATCTGAAGGTGGTAG GT AAGAATTCTAGAT
cDNA	gatctgaaggtg...
Genomic sequence	GATCTGAAGGTGGT <u>AAG</u> TAAGAATTCTAGAT

The *BTNL2* gene belongs to the butyrophilin family. Butyrophilin genes are members of the immunoglobulin superfamily^{143, 147}. According to sequence and structure homology, the entire butyrophilin family shares similarity with the B7 proteins¹⁴⁷ which are important in the regulation of previously activated T-cells¹⁴⁷⁻¹⁵⁰. Structurally, the *BTNL2* gene is organised into domains marked by the separate exons¹⁴⁴. The revised *BTNL2* protein consists of an N-terminal signal peptide encoded by exon 1 and 2, two N-terminal homologous IgV (immunoglobuline-like variable) domains encoded by exons 3 and 5, an heptad domain in exon 4 and a C-terminal IgC (immunoglobuline-like constant) domain encoded by exon 6. In the resulting protein model, a truncation of the protein at the site of the premature stop is introduced through rs2076530, just before the start of the IgC domain and the attached transmembrane helix. The variants P285L, P299Q and M286I are located in β -strands on the protein surface of the IgC domain (for details see appendix 9.2.4. and 9.2.5.).

3.2. Results of the *CARD15* gene analysis

3.2.1. *CARD15* mutations in sarcoidosis

Allele and genotype frequencies for all four *CARD15* SNPs in cases from the families and trios samples were compared to controls using chi-squared statistics. One patient per family was randomly drawn from the sarcoidosis families for this case-control analysis. Related patients were excluded to avoid over-representation of specific alleles from multi-case families. In addition to the population-based analysis, family-based analyses were performed by evaluating transmission disequilibrium. In total, 1011 individuals from the sarcoidosis families, the sarcoidosis trios and controls were genotyped for four Crohn's disease associated *CARD15* polymorphisms: SNP5, SNP8, SNP12, and SNP13. All genotyped SNPs were consistent with expectations of the Hardy-Weinberg law in controls.

Three different types of data analysis: case-control comparisons, genetic linkage analyses and family based association studies were performed to investigate the influence of *CARD15* mutations in sarcoidosis families and trios. In order not to overlook a possible effect, because of the relatively small sarcoidosis family and sarcoidosis trio sample sizes, both samples were also analysed together. Comparisons of allele frequencies for each SNP in 265 cases (138 cases from 138 sarcoidosis families, one patient at random chosen from each family, and 127 cases from 127 sarcoidosis trios) and 265 matched controls, revealed only two significant ($p < 0.05$) results. The frequency of allele 1 in both SNP5 and SNP12 was increased in cases. For SNP12 this effect was observed both in cases from sarcoidosis families and cases from sarcoidosis trios. However, only cases from sarcoidosis families showed a significant increase of SNP5 allele 1. Furthermore, a significant difference in genotype frequencies was observed only in SNP12 for the sarcoidosis trios and the combined cases from trios and families. All case-control comparisons are listed in table 13.

Table 13. *CARD15* allele (A) and genotype(B) frequencies in sarcoidosis patients and healthy controls. Numbering of *CARD15* mutations according to ref. ¹¹⁹. The rare allele of each SNP refers to allele 1.

A.

<i>CARD15</i> Alleles	<i>Controls</i> (n=265)	<i>Sarcoidosis families</i> (138 patients)	<i>Sarcoidosis trios</i> (127 patients)	<i>Sarcoidosis families + trios</i> (265 patients)
SNP5 1	144 (27,2%)	97 (35.1%)	80 (31.5%)	177 (33.4%)
SNP5 2	386 (72,8%)	179 (64.9%) p=0.019	174 (68.5%) p=0.21	353 (66.6%) p=0.027
SNP8 1	25 (4.7%)	15 (5.4%)	19 (7.5%)	34 (6.4%)
SNP8 2	505 (95.3%)	261 (94.6%) p=0.66	235 (92.5%) p=0.12	496 (93.6%) p=0.23
SNP12 1	4 (0.8%)	8 (2.9%)	8 (3.2%)	16 (3.0%)
SNP12 2	526 (99.2%)	268 (97.1%) p=0.017	246 (96.8%) p=0.011	514 (97.0%) p=0.007
SNP13 1	19 (3.6%)	11 (4.0%)	8 (3.2%)	19 (3.6%)
SNP13 2	511 (96.4%)	265 (96.0%) p=0.78	246 (96.8%) p=0.75	511 (96.4%) p=0.99

B.

<i>CARD15</i> genotypes	<i>Controls</i> (n = 265)	<i>Sarcoidosis families</i> (138 patients)	<i>Sarcoidosis trios</i> (127 patients)	<i>Sarcoidosis families + trios</i> (265 patients)
SNP5 1 / 1	18 (6.8%)	17 (12.3%)	11 (8.7%)	28 (10.6%)
SNP5 1 / 2	108 (40.7%)	63 (45.7%)	58 (45.7%)	121 (45.7%)
SNP5 2 / 2	139 (52.5%)	58 (42%) p=0.06	58 (45.7%) p=0.43	116 (43.8%) p=0.083
SNP8 1 / 1	0	0	1 (0.8%)	1 (0.4%)
SNP8 1 / 2	25(9.4%)	15 (10.9%)	17 (13.4%)	32 (12.1%)
SNP8 2 / 2	240 (90.6%)	123 (89.1%) p=0.65	109 (85.8%) p=0.17	232 (87.5%) p=0.37
SNP12 1 / 1	0	1 (0.7%)	0	1 (0.4%)
SNP12 1 / 2	4(1.5%)	6 (4.3%)	8 (6.3%)	14 (5.3%)
SNP12 2 / 2	261 (98.5%)	131 (95.0%) p=0.08	119 (93.7%) p=0.01	250 (94.3%) p=0.034
SNP13 1 / 1	0	0	0	0
SNP13 1 / 2	19 (7.2%)	11 (8.0%)	8 (6.3%)	19 (7.2%)
SNP13 2 / 2	246 (92.8%)	127 (92.0%) p=0.77	119 (93.7%) p=0.75	246 (92.8%) p=0.99

Family based association studies were performed to determine if there was any deviation from random expectation in the transmission of SNP alleles from heterozygote parents to affected offspring using the transmission disequilibrium test. 109 complete trios were extracted from the sarcoidosis families sample in addition to the 127 trios of the sarcoidosis trios set. No transmission disequilibrium was observed in trios taken from the sarcoidosis families sample, neither for single SNPs nor for 4 marker haplotypes. By contrast, significant increased transmission of SNP5 allele 1 as well as significantly decreased transmission of SNP allele 2 was detected in the sarcoidosis trios sample. Allele 1 was more often transmitted from heterozygote parents to affected offspring. Haplotype analysis demonstrated the same results, including significantly decreased transmission of haplotypes carrying SNP5 allele 2. Results of transmission disequilibrium tests are listed in table 14.

Table 14. Results of transmission disequilibrium test. The analysis use 109 complete trios extracted from the sarcoidosis families sample and 127 trios of the sarcoidosis trios set. Only the results of the rarer allele 1 of individual single nucleotide polymorphisms (SNP) are listed. For allele 2 of the biallelic polymorphisms, the number of transmitted and untransmitted alleles (denoted as trans and untrans respectively) is inverse, with identical P values. Haplotypes are formed by alleles of SNP5-SNP8-SNP12-SNP13. P - P value.

SNPs	Families			Trios			Families and trios		
	trans	untrans	p	trans	untrans	p	trans	untrans	P
SNP5 1	31	33	0.8	60	39	0.034	91	72	0.13
SNP8 1	8	8		18	10	0.13	26	18	0.23
SNP12 1	5	5		8	3	0.13	13	8	0.28
SNP13 1	8	7	0.8	8	8		16	15	0.86
Haplotypes									
1 1 2 2	8	7	0.8	14	7	0.13	22	14	0.18
1 2 1 2	5	4	0.74	6	2	0.16	11	6	0.23
1 2 2 1	3	6	0.32	8	6	0.59	11	12	0.83
1 2 2 2	18	19	0.87	31	23	0.28	49	42	0.46
2 1 2 2	0	0		1	3	0.32	1	3	0.32
2 2 1 2	0	0		0	0		0	0	
2 2 2 1	4	1	0.18	0	0		4	1	0.18
2 2 2 2	25	26	0.89	33	52	0.039	58	78	0.086

3.2.2. Sequence analysis of the CARD15 gene NOD domain

In 39 patients (from 39 families) sharing at least one parental chromosome 16 (identical by descent) with their affected sibling(s), none of the reported Blau syndrome mutations and no new mutations of exon 4 were detected. No new mutations were found in eight patients representing 7 complex families and sharing no parental chromosome 16. Sequence analysis of these 47 patients revealed only the known sequence variations denoted SNP5, SNP6, and SNP7 in ref ¹¹⁹. The frequencies of these alleles in 47 sequenced individuals did not differ from the frequencies reported for larger samples of healthy individuals of German origin ¹¹⁰.

4. Discussion

4.1. Association of the *BTNL2* mutations with sarcoidosis

4.1.1. *Experimental design*

A three-stage SNP mapping experiment with mutation detection in the linkage region was performed on chromosome 6p21 in order to localise genetic variants that may generate a predisposition to sarcoidosis. Because allele/genotype frequencies depend heavily on the genetic background and history of a population, families, trios and incomplete pedigrees were collected in addition to single affected cases so that both population- and family-based association study designs could be implemented. Such an experimental strategy, with both family and case-control analyses, was chosen so that the case-control studies, which can be prone to false positive (or false negative) results due to possible population admixture/sub-structure, could be validated. For population based analyses, unrelated controls were matched to cases according to ethnic background, age and gender ratio. In the family-based analyses, families with both parents were preferentially collected (if available) to obtain the maximum number of informative trios for the transmission disequilibrium tests.

A high throughput SNP typing method was applied to the association mapping experiment. It is an enabling technology that is useful for the identification of susceptibility genes by linkage disequilibrium mapping in large samples like in our study^{140, 151}. The method utilizes the 5' exonuclease activity of Taq polymerase and employs a combination of PCR and competitive hybridization¹⁵². The TaqMan genotyping results, as well as pedigree, patient phenotype, plate layout and marker information were stored and analysed in an integrated database system, which enabled to achieve the required throughput and to ensure the quality of the data¹³⁰.

4.1.2. *Identification of the association region on chromosome 6p*

The genotyping strategy involved screening the linkage region, fine mapping in the resulting association region and then verifying the disease-marker association in an independent sample set. The linkage region had previously been identified in a study of 55 sarcoidosis families⁶¹ and was further supported in an extended sample set containing 63 families⁵⁸. The linkage

was consistent with numerous reports on associations between gene variants in the region and the risk of sarcoidosis. The linkage region was screened with 69 SNP marker in a combined sample set of families, trios and single patients containing 265 independent cases in total (referred as „exploratory’’). Three marker haplotypes in case-control and transmission disequilibrium tests were analyzed to maximize the informativity of the SNPs. Only the *BTNL2-HLA-DRA* locus showed significance and demonstrated consistent results in both analyses. The association region peak was located at the *HLA-DRA* marker rs7192 ($P_{\text{TDT(global)}}=0.00068$; $P_{\text{case-control(LR)}}=0.00008$), decreasing at *BTNL2* marker rs2076523 ($P_{\text{TDT(global)}}=0.02512$ $P_{\text{case-control(LR)}}=0.00005$). Another region located around the *MICB-IkBL* genes was significant only in the case-control analysis (rs3134900 $P_{\text{case-control(LR)}}=0.00011$. $P_{\text{TDT(global)}}=0.14103$). This locus also demonstrated long range LD with the markers from the previous region, exhibiting D' values as large as 0.94. This indicates that the significance in the case-control analyses at *MICB-IkBL* may be influenced by the haplotype structure of the MHC locus. To resolve this problem, a genotype-based logistic regression analysis was performed that included markers from both regions. This analysis excluded the *MICB-IkBL* region and indicated that the disease susceptibility locus maps to the *BTNL2-HLA-DRA* locus. Consequently, the *BTNL2-HLA-DRA* region was selected for further fine mapping experiments. In order to verify the significant findings from the first screening experiment, and to test the validity of final mapping experiment result, an additional sample set containing trios and single cases (referred as the „extension’’ sample) was added to the „exploratory’’ sample. The resulting set of patients contained 495 independent cases, representing a larger population affected by sarcoidosis.

4.1.3. Sarcoidosis candidate genes on chromosome 6p

The establishment of a high density LD map in the association region involved a detailed analysis of the sequence in the on-line databases. The chromosomal sequence was analysed for its gene content, the function of the genes and possible functional mutations. At the beginning of this study, there were five known genes in the region covered by the SNP screening experiment. The *AGER* gene (located at 32.22Mb of chromosome 6) belongs to the immunoglobulin superfamily of cell-surface molecules and is an important receptor for the amyloid beta peptide. Expression of this receptor has been found to be increased in Alzheimer’s disease¹⁵³. The gene is also involved in inflammatory processes¹⁵⁴. The product of the *NOTCH4* gene (at 32.25Mb) functions as a receptor for membrane-bound ligands and it

may play a role in vascular, renal and hepatic development. This gene has been found to be associated with susceptibility to schizophrenia in a fraction of cases^{155, 156}. Two genes *TSBP* (at 32.37Mb) and *BTNL2* (at 32.44Mb) had unknown function. However, it was known, that *BTNL2* contains immunoglobulin domains¹⁴⁴ and is related to the B7-CD28 superfamily: these molecules have a crucial role in regulating T-cell activation and tolerance¹⁴⁷. Centromeric to the *BTNL2* gene lies in the *HLA* gene cluster. These play a central role in the immune system by presenting peptides derived from extracellular proteins¹⁷. Although many investigators have reported associations between *HLA-DRB1* and sarcoidosis, *BTNL2* became an interesting positional and functional candidate gene. The 440kb region bordered by the *AGER* and *HLA-DQA1* genes and containing the *BTNL2* gene therefore represented a prime-target for the establishment of a high-density SNP map. As *HLA* genes are highly polymorphic and TaqMan assay design is largely not possible in the region, the high-density SNP map covered *HLA-DRA* gene and finished at the 3-prime end *HLA-DRA* gene.

Every exon of the positional candidate gene *BTNL2* was sequenced in 47 sarcoidosis patients in order to verify the known sequence variations and to detect any not previously described mutations. 48 SNP markers were genotyped in the combined „exploratory” and „extension” sample set that contained 495 independent cases. As in the first screening experiment, both family and population based analyses were carried out on this data. A 15kb region at the 3’ end of the *BTNL2* gene, containing the markers rs2076530, P285L, M286I and P299Q, demonstrated a strong and consistent signal in both case-control and TDT analyses.

It is important that an association study finding should be replicable in independent sample sets. In this study, the initial association study results were verified in an independent sample set (referred to in our study as „verification”). For this, an additional sarcoidosis sample consisting mostly of trios and single patients, and yielding 452 independent cases in total, was genotyped at the four disease-associated markers: rs2076530, P285L, M286I and P299Q. The SNP (rs2076530), in the downstream region of the *BTNL2* gene, was most strongly associated with disease and this observation was replicated in the verification sample. This variant is common and has a minor allele (allele 2 that corresponds to nucleotide G) frequency of 0.42 in the control population. The allele 1 (corresponds to nucleotide A) apparently contributes to disease susceptibility in a manner that is completely independently of the (known) disease-associated alleles in the *HLA-DRB1* gene. The susceptibility allele A of rs2076530 is characterised by an odds ratio of 1.55 in heterozygotes, and of 2.56 in homozygotes,

corresponding to PARs of 22% and 34%, respectively. Comparing this result to other complex diseases associated loci, for example *CARD15* in Crohn disease, rs2076530 has only a moderate impact on an individual's disease risk. This is particularly intriguing in light of its profound impact upon the *BTNL2* protein structure and function. This apparent discrepancy may be in part explained by a highly redundant system of co-stimulatory molecules, including the products of the butyrophilin gene cluster, B7-1, plus other, as yet unknown, members of the same functional class of proteins.

The revised *BTNL2* transcript consists of eight exons. Experiments on cDNA indicated that the annotated *BTNL2* gene transcript was incomplete in that it was missing one exon at the 5' and two exons at the 3' end. Exon 3, referred to as an IgC domain in the original reference¹⁴⁴, was not identified in our transcripts, thereby rendering the hypothesis of two transcripts at this locus somewhat unlikely. Furthermore, the most significant variation, rs2076530, does not result in an amino acid exchange but rather appears to influence splicing through the replacement of the normally conserved G allele at the -1 donor site of exon 5 with an A.

The *BTNL2* expression experiments supported the hypothesis, that the gene is implicated in susceptibility to sarcoidosis. Expression was found only in the bronchoalveolar lavage of affected lungs and in myelomonocytic cell-lines stimulated with TNF α . It is known, that activated alveolar macrophages secrete increased amounts of TNF α in sarcoidosis patients⁹⁸. Therefore, it may be speculated that TNF α induces *BTNL2* transcription in the sarcoid lung.

Phylogenetic analyses and homology searches with the immunoglobulin domains of *BTNL2* indicated, that the gene is similar to five human butyrophilin genes and belongs to the extended butyrophilin family¹⁴⁴. *BTNL2* also shows similarity to mouse *NG9* and *NG10* genes¹⁴⁴ and is homologous to the Zipper proteins, bg2, MOG, and B-G which, phylogenetically, appear to be members of extended B7 family¹⁵⁷. The immunopathogenesis of demyelination in multiple sclerosis involves an autoantibody response to the immunoglobulin superfamily member myelin oligodendrocyte glycoprotein (MOG)¹⁵⁸. Crystal structure analysis has demonstrated that the extracellular domain of MOG adopts an IgV like fold that harbours a cavity that is similar to the one used by the co-stimulatory molecule B7-2 to bind its ligand CTLA4¹⁵⁹. Based on amino acid homology and domain structure, *BTNL2* is most similar to the B7-1 protein (for details see chapters 9.2.4. and 9.2.5.). B7-1 (CD80) is a co-stimulatory molecule with known anti-inflammatory activity that is mediated through an interaction with

CTLA-4^{160, 161}. Co-stimulatory pathways for T-cells are initiated when CTLA-4 and CD28 engage B7 class molecules in the context of a primary signal delivered through the TCR-MHC interaction^{149, 162, 163}. Deletion the IgC domain in the CD80 gene had a substantial pro-inflammatory effect in a mouse plasmid vaccination model, suggesting that the constant-like domains are critically important for immune regulation during co-stimulation¹⁶⁴. It may be therefore hypothesised that *BTNL2* potentially has a down-regulatory effect on T-cells and that this would be impaired by the truncating mutation at rs2076530. This leads to a loss of the IgC domain, thereby leading to inappropriate T-cell activation. This hypothesis would also be compatible with the clinical immunology of sarcoidosis, which is characterized by a dysregulated T-helper cell activation¹⁶⁵.

The importance of co-stimulatory dysfunction is becoming increasingly recognized in many autoimmune disorders¹⁶⁶. Our findings also emphasize the importance of the costimulatory mechanism in the regulation of T-cell activity, since CTLA4, a prototype of a putative *BTNL2* receptor, is a risk factor for such autoimmune disorders as type I diabetes, autoimmune hypothyroidism and Grave's disease¹⁶⁶. The exact function of *BTNL2* and its role in the costimulatory system remains to be elucidated. The identification of the *BTNL2* gene as a risk factor for sarcoidosis will give new impetus to the understanding of immunological co-stimulation. Functional studies of *BTNL2* may provide a perspective for future therapeutic targeting of this disorder.

4.1.4. Genotyping of the HLA-DRB1

Linkage to the *HLA* region has been shown for many complex autoimmune diseases. Despite the primary function of *HLA* genes in the immune system, many other immune response related genes also map to the region and their role in disease aetiology can not be excluded. However, only a few non-*HLA* disease genes have been successfully implicated with disease in the MHC region^{89, 90}. In the search for disease genes in this region, *HLA* genotyping is essential either to differentiate the gene effect from a nearby *HLA* gene or to implicate the *HLA* molecule itself in the disease aetiology. However, in such situations, high-throughput methods are required and low-resolution is sufficient. Current *HLA* genotyping methods do not meet such requirements.

In this study, a method was developed that combined allele-specific PCR with non-electrophoretic scoring of amplification products; an approach that is useful for high-throughput *HLA-DRB1* allele-group genotyping. Seven, non-overlapping groups of *HLA-DRB1* alleles were genotyped in place of the actual alleles^{127, 128}. Because *HLA-DRB1* is a multi-allelic system, all groups had to be typed for an individual in order to reconstruct the final genotype. The scoring of PCR amplification products usually involves agarose gel electrophoresis. However, this method does not scale well for high-throughput. A non-electrophoretic detection method was therefore adapted¹²⁵. This method specifically measures dsDNA concentration and detects an increase in samples where the PCR targeted *DRB1* allele is present. A specialized software was developed for the scoring of PCR amplification products, genotype assignment, Mendelian segregation checking and integration into the laboratory database system. This allowed flexible handling of the *DRB1* allele-group genotype information.

In general, multi-allelic systems demonstrate a higher rate of genotyping error when compared to biallelic SNP genotyping. An individual can be miscalled as a homozygote simply because of a single failed PCR reaction. In our method, DNA concentration is further critical parameter because the detection of PCR amplification products with the dsDNA-specific dye PicoGreen® is based on an increase in the dsDNA concentration when the targeted allele is present. To achieve uniform and reproducible genotyping results, the initial DNA concentration was measured and adjusted to 2.5ng/well for all analysed samples. The approach was evaluated in 2,295 individuals and gave a genotyping success rate of 94.11%. Mendel error rates in the family sample (1.3%) compared favourably with other genotyping methodologies¹³¹.

The primary purpose of this study was to establish a high-throughput *HLA-DRB1* allele-group typing method for disease gene finding in the MHC region and to facilitate the differentiation of the genetic effect of a disease gene from a nearby *HLA* gene. Nevertheless, the *HLA-DRB1* allele genotyping results were also compared with associations reported in the literature for other populations. In principle, the results obtained in the current study were in agreement with previously described associations in distinct Caucasian ethnic groups. DRB1*01 and DRB1*04 were reported as „protective” alleles in case-control studies with UK, Czech, Polish, Scandinavian, Italian and US Caucasian patients (with the exception of DRB1*01 in the US Caucasians)^{72, 74}. Although the frequencies of the “protective” *HLA* alleles varied to

some extent among the control samples of these studies, the reduction in allele frequency in patients was consistent across all investigated populations ⁷². In the current study design, the DRB1*01 and DRB1*04 alleles were amplified in allele amplification groups 1 and 4 respectively. For both these allele groups reduced allele frequencies in samples of affected individuals were detected. As expected, this result was in agreement with the pattern of observed and expected transmissions in TDT analyses. That *DRB* alleles carrying hydrophobic side-chain residues at position 11 of the HLA-DRB1 peptide may confer a protective effect remains open to exploration. As the disease causing antigenic stimulus is unknown, it is still unclear whether the possible protective effect of *HLA-DRB1* alleles could be due to an altered binding of autoantigens or foreign antigens ⁷².

In contrast to the consistent reports of „protective” HLA-DR alleles across a variety of populations, descriptions of alleles associated with sarcoidosis susceptibility demonstrate non-overlapping results, and seem to appear to act in ethnicity-dependent manner, or to influence the course of disease. For example, the alleles carrying hydrophilic residues at position 11 (HLA-DRB1*07, *08, *09) apparently influence disease susceptibility in different ways in UK, Polish, Czech, Scandinavian and Japanese patients. HLA-DR-14 and -15 were found to be more common in UK, Polish and Czech populations ⁷². HLA-DR17 was found to be particularly associated with an acute disease course in Scandinavian and German cases ^{70, 71}. HLA-DRB1*1101 has been suggested to be the most important susceptibility factor for sarcoidosis among the HLA class II alleles on the grounds that this allele has been significantly associated with sarcoidosis in both people of African ancestry and Caucasians ⁷⁴. In the current study, amplification group 3 alleles (DRB1*03, *08, *11(not *1122/30, *12, *13, *14, (not *1410/39)) appeared confer an increased risk of sarcoidosis. The frequency of group 3 alleles was significantly increased in sarcoidosis cases relative to controls. A significant over-transmissions of these alleles was also observed in the family sample, thus supporting the case-control finding. It was therefore possible that *HLA* alleles and not the sequence variation at rs2076530 in *BTNL2* might be driving the association effect in the region and that the signal at rs2076530 resulted from linkage disequilibrium between the markers. However, separation of the gene effects via a stratification analysis of the respective *HLA-DRB1* and *BTNL2* genotypes using a Breslow-Day test for homogeneity of odds ratios has rejected this hypothesis. Therefore, it appears that *HLA-DRB1* amplification group 3 and rs2076530 allele A represent independent risk factors for sarcoidosis.

4.2. CARD15 mutations in sarcoidosis

Sarcoidosis and Crohn's disease are multifactorial disorders, in which disease is caused by an inherited susceptibility and an additive influence of unknown exogenous agents^{33, 34, 59}. In contrast, Blau syndrome is caused by single dominant gene mutation that directly leads to Mendelian inheritance of the disorder, although yet unknown factors modify the expression of the disorder between and within affected families considerably^{60, 104}. The clear mode of monogenic inheritance helped to localise the Blau syndrome gene. It was mapped to the centromeric region of chromosome 16 by linkage analysis of only one extended family¹⁰⁴. In contrast to monogenic diseases, large numbers of affected families and repeated genome-wide linkage studies are necessary to locate disease genes in complex diseases¹⁰⁵⁻¹⁰⁷. Such a strategy finally led to the successful identification of CARD15 mutations as a major contributory factor in Crohn's disease susceptibility¹⁰⁸⁻¹¹⁰. In sarcoidosis, genome-wide linkage information is so far limited to one report⁵⁸ showing a main peak at the MHC region on the short arm of chromosome 6^{61, 167}. The genetic mechanisms behind a predisposition to sarcoidosis are unclear, although the contribution and cooperation of several genetic factors in the development of sarcoidosis is likely. Therefore the peak at the MHC region cannot fully explain the increased familial recurrence risk found in different studies in European populations^{42, 168, 169} and in the USA⁵⁴.

In addition to a genome-wide search by linkage analysis in affected siblings, association studies with candidate genes are widely used for the study of complex diseases²⁸. However, heterogeneity in sarcoidosis aetiology is likely to be caused by variable influences from different genetic factors that depend on the ethnic background. This could be the reason for the many contrasting reports on associations between sarcoidosis and polymorphisms in candidate genes⁶¹. Nevertheless, there is an on-going search for promising candidate genes that connects an inhaled microbial trigger to a balanced cellular immune response⁷⁷.

Significant familial clustering of sarcoidosis and Crohn's disease has been reported.⁴² This suggests homologous mechanisms in the pathophysiology of both disorders. In Crohn's disease, increased levels of NF kappa B play an important role in the development and maintenance of inflammation¹¹¹. Increased levels of NF kappa B have been also found in sarcoidosis¹¹². Since the function of the CARD15 protein includes the regulation of NF kappa B levels, mutations in the *CARD15* gene in patients with Crohn's disease seem to clarify an

important link in the pathway from a presumed exogenous agent to granulomatous inflammation process.

Blau syndrome is a rare condition with a limited number of reports in literature. The first mutations were identified in four unrelated families ¹³². All patients from these families carried a mutation in the nucleotide oligomerisation domain (NOD) of the *CARD15* gene. In three families the same amino acid of the NOD, arginine at amino acid position 334 of the protein, was substituted by glutamine or tryptophan (R334Q and R334W). Both mutations were confirmed in another 5 out of 10 unrelated families ¹³³. According to these findings, Blau syndrome mutations appear to cluster strongly within the NOD, and mutations of the LRR seem to be linked with Crohn's disease. Thus the mutations in different domains of the *CARD15* protein cause different phenotypes within the spectrum of granulomatous disorders, leading to the development of Blau syndrome or to a significant increase in the risk of Crohn's disease, respectively. Taking into account the considerable overlap of clinical features between Crohn's disease, Blau syndrome and sarcoidosis, the *CARD15* gene could be considered an excellent candidate gene of sarcoidosis susceptibility.

There was preliminary information concerning the Blau syndrome mutation region in the *CARD15* gene of sarcoidosis patients ¹⁷⁰. In a study of 33 African-American nuclear families with 67 sarcoidosis siblings, tested for co-segregation of the disease with eight highly-polymorphic DNA markers from the centromeric region of chromosome 16, no evidence of linkage was found. However, the heterogeneity of disease between African-American and Caucasians ^{33, 55} suggests that the results from this study may not necessarily be valid in Caucasians.

A subgroup of the study population used in this experiment has been analysed in a genome-wide linkage study ⁵⁸. The microsatellite marker D16S3396 analysed in the scan is located close to the *CARD15* gene. The results did not provide any significant evidence ($P = 0.21$) that siblings suffering from sarcoidosis tend to be identical by descent for this region of chromosome 16. However, these linkage data do not rule out the possibility that *CARD15* mutations could play a role in a subgroup of sarcoidosis patients. Therefore, mutations of the *CARD15* gene that could contribute to sarcoidosis in only a portion of families seemed likely. In the current study, we have analysed the sequence of the *CARD15* gene region that is altered in Blau syndrome, in a selected panel of patients. We excluded the presence of Blau syndrome

mutations in sarcoidosis patients and detected no other sequence variations than the known polymorphic sites. This result from sarcoidosis families cannot exclude a minor influence of rare NOD sequence variants in sarcoidosis. However, it clearly implies that a major contribution of CARD15 NOD variants in Caucasian patients is unlikely because a major gene effect would be linked to the familial occurrence of sarcoidosis.

In order to investigate a minor contribution of Crohn's disease CARD15 mutations to the aetiology of sarcoidosis, extended panels of sarcoidosis families, sarcoidosis trios and controls were genotyped for four SNPs that are associated with an increased risk of Crohn's disease. With the exception of SNP13 in the sarcoidosis families set, the rarer alleles of the four SNPs tended to be more frequent in cases than in controls, replicating the observations made in Crohn's disease study¹⁰⁸⁻¹¹⁰. A significant difference was detected for SNP5 in sarcoidosis families and for SNP12 in sarcoidosis families and in sarcoidosis trios. The allele frequencies of SNP5, SNP8 and SNP12 in our controls were in agreement with the published frequencies for controls in an Europe-wide study of Crohn's disease¹¹⁹, however, the SNP13 minor allele frequency of our study was considerably higher (3.6% versus 2%). On the other hand, a German control sample created for the case-control study of Crohn's disease had an even higher frequency of SNP13 allele 1 of 5.3%¹¹⁰. Genotyping of our extended group of available German controls, who were excluded from this study because of young age, indicates that the SNP12 allele 1 frequency in the matched control group is relatively low. Therefore, significant differences between sarcoidosis cases and controls might be attributable in part to this control group related phenomenon.

Taking into account the limitations of case-control comparisons in association studies of candidate gene polymorphisms in complex disorders, family-based strategies such as the TDT have been employed in this study to overcome control group related problems. Although sarcoidosis trios exhibited an over-transmission of SNP5 allele 1, this observation did not overlap with the case-control results for this marker in the same set of patients. In addition, further information about haplotype segregation extracted from families has been analysed. Strong linkage disequilibrium between the four SNPs has been observed in other studies^{26, 171}, and can provisionally be derived from the haplotype proportions of the TDT analysis. As listed in table 14, the rare alleles of SNP8, SNP12 and SNP13 occur almost exclusively in combination with SNP5 allele 1.

In the trios, all haplotypes with the rare allele of at least one of the four SNPs are preferentially transmitted from heterozygote parents to affected offspring although this deviation does not reach significance. However, complementary haplotypes composed of the four common alleles demonstrates a significant under-transmission. Interestingly, transmission of alleles and haplotypes is almost perfectly balanced in the sarcoidosis families. This fact could point to heterogeneity in the genetic aetiology of sarcoidosis with respect to *CARD15* gene variation.

5. Conclusions

The chromosome 6 association study in this thesis uses an association mapping method for identification of a disease susceptibility gene. The experiment design included both family- and population-based study approaches. The basic criterion for the validity of association findings was consistency of results in both family- and population based study designs. Using a single nucleotide polymorphism typing method, 16.4 Mb of the chromosome 6p21 region that was implicated in the chromosome 6 linkage study⁵⁸ were scanned in order to identify primary association lead(s). Consistently significant results in both study designs were obtained for one peak located at the *BTNL2-HLA-DRA* region. To map the disease-associated gene, a certain marker density had to be achieved in the lead region. A high-density SNP map in the primary association region was therefore established through a detailed analysis of the physical map and the identification of sequence variations. Using TDT and case-control analysis, a 15kb segment of the *BTNL2* gene was identified as being strongly associated with sarcoidosis. The major disease-associated variant rs2076530, located in exon 5 of *BTNL2* represents a risk factor that is entirely independent of the previously reported association between sarcoidosis and alleles of the *DRB1* gene. The risk allele A of rs2076530 leads to alternative splicing of the *BTNL2* transcript, which introduces a premature stop codon and leads to the truncation of the protein. *BTNL2* is a member of the immunoglobulin superfamily and has been implicated as a co-stimulatory molecule based on its homology to B7-1.

The *BTNL2* gene, containing the disease susceptibility allele (rs2076530), is located in the close vicinity (~200kb) of the *HLA-DRB1* gene. Since the *HLA* gene has also been implicated in the aetiology of disease, a special method of allele group-specific PCR typing with a non-electrophoretic method and novel software to support scoring have been developed for high-throughput typing. By design, primers amplified non-overlapping groups of *HLA-DRB1* alleles and only limited resolution of *DRB1* alleles was achieved. Separate PCR-reactions had to be performed for each allele group in order to construct the final genotype. The method has several advantages over the alternatives and will complement other approaches^{126, 172, 173}. dsDNA measurement was used instead of agarose gel electrophoresis that does not scale well for high-throughput applications. Measurement of dsDNA is much less expensive than TaqMan assays¹⁷⁴ or fluorescently labeled PCR products for automated sequencing. The low-resolution *HLA-DRB1* allele genotyping strategy developed in this study has been

successfully applied for 2,295 individuals. The method may facilitate the gene identification of complex, *HLA* linked autoimmune disorders.

A candidate gene approach is a useful approach for association testing between a disease and positional or functional candidate genes. In this study, *CARD15*, which leads to the development of Blau syndrome or to a significant increase in the risk of Crohn's disease has been selected as a sarcoidosis candidate gene. In a case-control analysis, SNP12 alele1 was more frequent in sarcoidosis cases comparing to the controls. From the TDT results, sarcoidosis trios but not families demonstrate significant result for SNP5. Thus, replication of the investigation is desirable. Nevertheless, the present association and sequence analysis results seem to rule out a major role of *CARD15* mutations in the pathogenesis of sarcoidosis. Because Crohn's disease, Blau syndrome and sarcoidosis significantly overlap with respect to their pathophysiology and clinical phenotypes, this finding demonstrates the genetic heterogeneity of granulomatous inflammatory disorders.

6. Summary: Identification of sarcoidosis susceptibility genes by association mapping and candidate gene approaches

Sarcoidosis is a systemic immune disorder with predominant manifestations in the lung, lymph nodes, eyes and skin. The disease can lead to pulmonary fibrosis and chronic respiratory failure and is pathophysiologically characterized by an exaggerated T-helper cell response. Identification of a sarcoidosis susceptibility gene on chromosome 6 in this study has been performed using the association mapping approach. In a previous genome wide linkage analysis, the extended MHC region on chromosome 6p has been implicated in the aetiology of sarcoidosis. Here, a systematic three-stage single nucleotide polymorphism (SNP) scan of 16.4 Mb on chromosome 6p21 was performed in up to 947 independent cases of familial and sporadic sarcoidosis. Using TDT and case-control analyses, a 15 kb segment located at the 3 prime end of the *BTNL2* gene could be identified as being strongly associated with sarcoidosis. The major disease-associated variant, rs2076530, represents a risk factor that is entirely independent of the previously reported association between sarcoidosis and alleles of the *DRB1* gene, located within ~200kb of *BTNL2*. *BTNL2* is a member of the immunoglobulin superfamily. Homology to B7-1 implicates the *BTNL2* as a co-stimulatory molecule. The risk allele A of rs2076530 leads to alternative splicing of the *BTNL2* transcript, which introduces a premature stop. The resulting truncated protein lacks the C-terminal IgC domain and transmembrane helix, thereby disturbing the putative co-stimulatory function of this molecule.

Differentiating specific gene effects from the nearby *HLA* loci is a common problem in the mapping of susceptibility genes of disorders linked to the MHC region. The *HLA* molecules may be risk factors for the disease themselves or they may serve as markers tagging risk haplotypes. Since the *BTNL2* gene with the disease-associated variant is located near *HLA-DRB1*, which has itself previously been implicated as a risk factor for sarcoidosis, genotyping of the *HLA* was performed to evaluate the effects of both genes. In this situation, high-throughput rather than detailed resolution of the *HLA* genotype was required. Current methods do not support to this kind of requirement. A method combining non-electrophoretic scoring and genotyping of non-overlapping allele -groups, instead of specific *HLA-DRB1* alleles, has been developed for genotyping large patient samples. A specialized software facilitated the construction of the final *DRB1* genotype, testing of Mendelian segregation and database import. The approach was practically evaluated in a sample of 2,295 individuals from nuclear

families and 538 unrelated control individuals. Genotyping success rate was 94.11% and only 1.3% of the data had to be discarded due to Mendel errors. The *HLA-DRB1* allele genotyping strategy developed in this study may contribute to gene identification in other complex, *HLA* linked autoimmune disorders.

Association studies of candidate genes are widely used for the study of complex diseases. Crohn's disease, Blau syndrome and sarcoidosis significantly overlap in clinical phenotypes and pathophysiology. In our sarcoidosis study, a minor contribution of *CARD15* SNP12 cannot be excluded because allele 1 was significantly more frequent in cases than controls. TDT results indicated association in sarcoidosis trios but not in families, thus replication of the investigation is desirable. However, association and sequence analysis results rule out a major role for the *CARD15* mutations that confer susceptibility to Crohn's disease and Blau syndrome in the pathogenesis of sarcoidosis in Caucasians. This finding demonstrates the genetic heterogeneity of granulomatous inflammatory disorders.

The association mapping and candidate gene testing results in patients suffering from sarcoidosis should provide insights into the pathogenesis of this complex disease.

7. Zusammenfassung: Identifikation von Sarkoidose Krankheitsgenen durch Assoziationsstudien und Kandidatengen Ansätze

Sarkoidose ist eine systemische Immunkrankheit mit vorwiegenden Manifestationen in der Lunge, den Lymphknoten, den Augen und der Haut. Die Krankheit kann zu Lungenfibrose und chronischem Atemwegsversagen führen. Pathophysiologisch wird Sarkoidose durch eine stark übertriebene T-Helferzellen Antwort charakterisiert. Mit Hilfe von Assoziationsstudien konnte ein Sarkoidose Empfänglichkeitsgen auf Chromosom 6 identifiziert werden. Eine vorher durchgeführte, genomweite Kopplungsstudie belegt einen Zusammenhang zwischen der erweiterten MHC Region auf Chromosom 6p und der Ätiologie von Sarkoidose. Daraufhin wurde ein systematischer, dreistufiger Einzelnukleotid Polymorphismus (SNP) Scan der 16,4 Mb auf Chromosom 6p21 für 947 unabhängige Fälle von familiärer und spontaner Sarkoidose durchgeführt. Durch das Benutzen des TDT und von Patient-/Kontrollanalysen, konnte ein 15 kb grosses Segment in der Nähe des 3' Endes des BTNL2 Genes identifiziert werden, welches stark mit Sarkoidose assoziiert ist. Die am häufigsten mit der Krankheit assoziierte Variante, rs2076530, stellt einen Risikofaktor dar, welcher komplett unabhängig von der schon länger bekannten Assoziation zwischen Sarkoidose und DRB1 Allelen ist. DRB1 liegt ~200 kb entfernt von BTNL2. Letzteres ist ein Mitglied der Immunglobulin Superfamilie. Die Homologie zu B7-1 impliziert, dass BTNL2 als co-stimulierendes Molekül fungiert. Das Risiko-Allel A von rs2076530 führt zum alternativen Splicing des BTNL2 Transkriptes, welches zu einem frühzeitigen Abbruch der Translation führt. Dem resultierenden gekürzten Protein fehlt die C-terminale IgC-Domäne samt transmembraner Helix, weshalb die vermutete co-stimulierende Funktion des Moleküls gestört wird.

Es ist ein bekanntes Problem, spezifische Geneffekte von nahegelegenen HLA Loci zu unterscheiden, wenn man potentielle Krankheitsgene kartiert, welche mit der MHC Region in Verbindung stehen. Die HLA Moleküle können selber Risikofaktoren für die Krankheit darstellen bzw. Marker auf Risiko Haplotypen sein. Da das BTNL2 Gen mit der krankheitsassoziierten Variante in der Nähe von HLA-DRB1 liegt, welches zuvor schon als Risikofaktor für Sarkoidose identifiziert wurde, wurde die HLA Region genotypisiert, um die Effekte der beiden Gene abzuschätzen. An Stelle von detaillierter Auflösung des HLA Genotyps, wurde eine Hochdurchsatz Methode angewandt. Da aktuelle Methoden noch nicht die nötigen Anforderungen erfüllen, wurde eine Methode entwickelt, welche nicht-

elektrophoretisches scoring mit Genotypisieren von nicht-überlappenden Allel-Gruppen kombiniert, um eine hohe Anzahl an Patienten zu genotypisieren. Dabei erleichterte eine spezielle Software die Konstruktion des endgültigen DRB1 Genotyps, den Test auf Mendelsche Vererbung, als auch den Datenbank Import. Der Versuch wurde mit einer Anzahl von 2.295 Individuen aus Kernfamilien und 538 nicht miteinander verwandten Kontroll-Individuen durchgeführt. Die Erfolgsrate lag beim Genotypisieren bei 94,11%, und nur 1,3% der Daten mussten wegen auftretender Mendelfehler vernachlässigt werden. Die HLA-DRB1 Allel Genotypisierungs-Strategie, welche in dieser Studie entwickelt wurde, könnte für die Genidentifikation bei anderen komplexen, HLA gekoppelten Autoimmunkrankheiten von grosser Bedeutung sein.

Assoziationsstudien an Kandidatengen werden generell für Studien von komplexen Krankheiten genutzt. Die Krankheiten Morbus Crohn, Blau Syndrom und Sarkoidose überlappen sich signifikant bezüglich klinischer Phänotypen und ihrer Pathophysiologie. So konnte in der Vergangenheit das Krankheitsgen CARD15 bei Morbus Crohn und beim Blau Syndrom identifiziert werden. In unser Sarkoidose Studie können wir nicht den geringen Einfluss von CARD15 SNP12 ausschliessen, da Allel 1 häufiger in Patienten als in Kontrollpersonen zu finden ist. Die Ergebnisse des TDT lassen auf eine Assoziation in Sarkoidose Trios aber nicht in Familien schliessen. Es wäre wünschenswert diese Ergebnisse durch erneute Studien zu festigen und zu verifizieren. Trotz allem schließen die CARD15 Assoziations- und Sequenzanalysen eine grundlegende Beteiligung an der Pathogenese der Sarkoidose aus. Diese Ergebnisse demonstrieren die genetische Heterogenität von granulomatösen entzündlichen Krankheiten.

Die Ergebnisse der Assoziationsstudie und die Kandidatengen Testergebnisse sollten einen Einblick in die Pathogenese der komplexen Krankheit Sarkoidose geben.

8. References

1. Morton NE. Sequential tests for the detection of linkage. *Am J Hum Genet* 1955; 7:277-318.
2. Nakamura Y, Leppert M, O'Connell P, et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 1987; 235:1616-22.
3. Suarez BK, Rice J, Reich T. The generalized sib pair IBD distribution: its use in the detection of linkage. *Ann Hum Genet* 1978; 42:87-94.
4. Schaid DJ, Nick TG. Sib-pair linkage tests for disease susceptibility loci: common tests vs. the asymptotically most powerful test. *Genet Epidemiol* 1990; 7:359-70.
5. Kruglyak L, Lander ES. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 1995; 57:439-54.
6. Gibbs RA, Belmont JW, Hardenbol P, et al. The International HapMap Project. *Nature* 2003; 426:789-96.
7. Goldstein DB. Islands of linkage disequilibrium. *Nat Genet* 2001; 29:109-11.
8. Lewontin RC. The Interaction of Selection and Linkage. Ii. Optimum Models. *Genetics* 1964; 50:757-82.
9. Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 1995; 29:311-22.
10. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001; 29:229-32.
11. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002; 296:2225-9.
12. Patil N, Berno AJ, Hinds DA, et al. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 2001; 294:1719-23.
13. Reich DE, Cargill M, Bolk S, et al. Linkage disequilibrium in the human genome. *Nature* 2001; 411:199-204.
14. Cullen M, Perfetto SP, Klitz W, Nelson G, Carrington M. High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *Am J Hum Genet* 2002; 71:759-76.
15. Chapman NH, Thompson EA. Linkage disequilibrium mapping: the role of population history, size, and structure. *Adv Genet* 2001; 42:413-37.
16. Johnson GC, Esposito L, Barratt BJ, et al. Haplotype tagging for the identification of common disease genes. *Nat Genet* 2001; 29:233-7.

17. Kumanovics A, Takada T, Lindahl KF. Genomic organization of the mammalian MHC. *Annu Rev Immunol* 2003; 21:629-57.
18. Nei M, Gu X, Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci U S A* 1997; 94:7799-806.
19. Ahmad T, Neville M, Marshall SE, et al. Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. *Hum Mol Genet* 2003; 12:647-56.
20. Jeffreys AJ, Kauppi L, Neumann R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* 2001; 29:217-22.
21. Walsh EC, Mather KA, Schaffner SF, et al. An integrated haplotype map of the human major histocompatibility complex. *Am J Hum Genet* 2003; 73:580-90.
22. Stenzel A, Lu T, Koch WA, et al. Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. *Hum Genet* 2004; 22:22.
23. Ziv E, Burchard EG. Human population structure and genetic association studies. *Pharmacogenomics* 2003; 4:431-41.
24. Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001; 2:91-9.
25. Spielman RS, Ewens WJ. The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 1996; 59:983-9.
26. Croucher PJ, Mascheretti S, Hampe J, et al. Haplotype structure and association to Crohn's disease of CARD15 mutations in two ethnically divergent populations. *Eur J Hum Genet* 2003; 11:6-16.
27. Iannuzzi MC, Maliarik M, Rybicki BA. Nomination of a candidate susceptibility gene in sarcoidosis: the complement receptor 1 gene. *Am J Respir Cell Mol Biol* 2002; 27:3-7.
28. Tabor HK, Risch NJ, Myers RM. Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 2002; 3:391-7.
29. Blencowe B. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci.* 2000; 25:106-10.
30. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 2002; 3:285-98.
31. Genin E. Selection of single nucleotide polymorphisms for association studies in candidate genes. *Genet Epidemiol* 2001; 21:S614-9.

32. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet* 2001; 29:306-9.
33. 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* 1999; 160:736-55.
34. Muller-Quernheim J. Sarcoidosis: immunopathogenetic concepts and their clinical application. *Eur Respir J* 1998; 12:716-38.
35. Hillerdal G, Nou E, Osterman K, Schmekel B. Sarcoidosis: epidemiology and prognosis. A 15-year European study. *Am Rev Respir Dis* 1984; 130:29-32.
36. Pietinalho A, Hiraga Y, Hosoda Y, Lofroos AB, Yamaguchi M, Selroos O. The frequency of sarcoidosis in Finland and Hokkaido, Japan. A comparative epidemiological study. *Sarcoidosis* 1995; 12:61-7.
37. Siltzbach LE, James DG, Neville E, et al. Course and prognosis of sarcoidosis around the world. *Am J Med* 1974; 57:847-52.
38. Hunninghake GW, Costabel U, Ando M, et al. ATS/ERS/WASOG statement on sarcoidosis. American Thoracic Society/European Respiratory Society/World Association of Sarcoidosis and other Granulomatous Disorders. *Sarcoidosis Vasc Diffuse Lung Dis* 1999; 16:149-73.
39. Drent M, Wirnsberger RM, Breteler MH, Kock LM, de Vries J, Wouters EF. Quality of life and depressive symptoms in patients suffering from sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* 1998; 15:59-66.
40. De Vries J, Drent M, Van Heck GL, Wouters EF. Quality of life in sarcoidosis: a comparison between members of a patient organisation and a random sample. *Sarcoidosis Vasc Diffuse Lung Dis* 1998; 15:183-8.
41. Wirnsberger RM, de Vries J, Breteler MH, van Heck GL, Wouters EF, Drent M. Evaluation of quality of life in sarcoidosis patients. *Respir Med* 1998; 92:750-6.
42. Wirnsberger RM, de Vries J, Wouters EF, Drent M. Clinical presentation of sarcoidosis in The Netherlands an epidemiological study. *Neth J Med* 1998; 53:53-60.
43. Iannuzzi MC. Clara cell protein in sarcoidosis: another job for the respiratory tract protector? *Am J Respir Crit Care Med* 2004; 169:143-4.
44. McGrath DS, Goh N, Foley PJ, du Bois RM. Sarcoidosis: genes and microbes--soil or seed? *Sarcoidosis Vasc Diffuse Lung Dis* 2001; 18:149-64.

45. du Bois RM, Goh N, McGrath D, Cullinan P. Is there a role for microorganisms in the pathogenesis of sarcoidosis? *J Intern Med* 2003; 253:4-17.
46. Hance AJ. The role of mycobacteria in the pathogenesis of sarcoidosis. *Semin Respir Infect* 1998; 13:197-205.
47. Jindal SK, Gupta D, Aggarwal AN. Sarcoidosis in developing countries. *Curr Opin Pulm Med* 2000; 6:448-54.
48. Maeda H, Niimi T, Sato S, et al. Human herpesvirus 8 is not associated with sarcoidosis in Japanese patients. *Chest* 2000; 118:923-7.
49. De Vuyst P, Dumortier P, Schandene L, Estenne M, Verhest A, Yernault JC. Sarcoidlike lung granulomatosis induced by aluminum dusts. *Am Rev Respir Dis* 1987; 135:493-7.
50. Skelton HG, 3rd, Smith KJ, Johnson FB, Cooper CR, Tyler WF, Lupton GP. Zirconium granuloma resulting from an aluminum zirconium complex: a previously unrecognized agent in the development of hypersensitivity granulomas. *J Am Acad Dermatol* 1993; 28:874-6.
51. Demedts M, Wells AU, Anto JM, et al. Interstitial lung diseases: an epidemiological overview. *Eur Respir J Suppl* 2001; 32:2s-16s.
52. Rybicki BA, Major M, Popovich J, Jr., Maliarik MJ, Iannuzzi MC. Racial differences in sarcoidosis incidence: a 5-year study in a health maintenance organization. *Am J Epidemiol* 1997; 145:234-41.
53. Baughman RP, Teirstein AS, Judson MA, et al. Clinical characteristics of patients in a case control study of sarcoidosis. *Am J Respir Crit Care Med* 2001; 164:1885-9.
54. Rybicki BA, Iannuzzi MC, Frederick MM, et al. Familial aggregation of sarcoidosis. A case-control etiologic study of sarcoidosis (ACCESS). *Am J Respir Crit Care Med* 2001; 164:2085-91.
55. Rybicki BA, Kirkey KL, Major M, et al. Familial risk ratio of sarcoidosis in African-American sibs and parents. *Am J Epidemiol* 2001; 153:188-93.
56. Jorgensen G. The genetics of sarcoidosis. *Acta Med Scand Suppl* 1964; 425:209-12.
57. Willoughby JM, Mitchell DN, Wilson JD. Sarcoidosis and Crohn disease in siblings. *Am Rev Respir Dis* 1971; 104:249-54.
58. Schurmann M, Reichel P, Muller-Myhsok B, Schlaak M, Muller-Quernheim J, Schwinger E. Results from a genome-wide search for predisposing genes in sarcoidosis. *Am J Respir Crit Care Med* 2001; 164:840-6.
59. Podolsky DK. Inflammatory bowel disease (1). *N Engl J Med* 1991; 325:928-37.

60. Blau EB. Familial granulomatous arthritis, iritis, and rash. *J Pediatr* 1985; 107:689-93.
61. Schurmann M, Lympny PA, Reichel P, et al. Familial sarcoidosis is linked to the major histocompatibility complex region. *Am J Respir Crit Care Med* 2000; 162:861-4.
62. Rybicki BA, Maliarik MJ, Poisson LM, et al. The major histocompatibility complex gene region and sarcoidosis susceptibility in African Americans. *Am J Respir Crit Care Med* 2003; 167:444-9.
63. Cooke GS, Hill AV. Genetics of susceptibility to human infectious disease. *Nat Rev Genet* 2001; 2:967-77.
64. Martinetti M, Luisetti M, Cuccia M. HLA and sarcoidosis: new pathogenetic insights. *Sarcoidosis Vasc Diffuse Lung Dis* 2002; 19:83-95.
65. Lenhart K, Kolek V, Bartova A. HLA antigens associated with sarcoidosis. *Dis Markers* 1990; 8:23-9.
66. Ina Y, Takada K, Yamamoto M, Morishita M, Senda Y, Torii Y. HLA and sarcoidosis in the Japanese. *Chest* 1989; 95:1257-61.
67. Kunikane H, Abe S, Yamaguchi E, et al. Analysis of restriction fragment length polymorphism for the HLA-DR gene in Japanese patients with sarcoidosis. *Thorax* 1994; 49:573-6.
68. Kunikane H, Abe S, Tsuneta Y, et al. Role of HLA-DR antigens in Japanese patients with sarcoidosis. *Am Rev Respir Dis* 1987; 135:688-91.
69. Nowack D, Goebel KM. Genetic aspects of sarcoidosis. Class II histocompatibility antigens and a family study. *Arch Intern Med* 1987; 147:481-3.
70. Swider C, Schnittger L, Bogunia-Kubik K, et al. TNF-alpha and HLA-DR genotyping as potential prognostic markers in pulmonary sarcoidosis. *Eur Cytokine Netw* 1999; 10:143-6.
71. Berlin M, Fogdell-Hahn A, Olerup O, et al. HLA-DR predicts the prognosis in Scandinavian patients with pulmonary sarcoidosis
Myelin oligodendrocyte glycoprotein: a novel candidate autoantigen in multiple sclerosis. *Am J Respir Crit Care Med* 1997; 156:1601-5.
72. Foley PJ, McGrath DS, Puscinska E, et al. Human leukocyte antigen-DRB1 position 11 residues are a common protective marker for sarcoidosis. *Am J Respir Cell Mol Biol* 2001; 25:272-7.

73. Sharma SK, Balamurugan A, Pandey RM, Saha PK, Mehra NK. Human leukocyte antigen-DR alleles influence the clinical course of pulmonary sarcoidosis in Asian Indians. *Am J Respir Cell Mol Biol* 2003; 29:225-31.
74. Rossman MD, Thompson B, Frederick M, et al. HLA-DRB1*1101: a significant risk factor for sarcoidosis in blacks and whites. *Am J Hum Genet* 2003; 73:720-35.
75. Foley PJ, Lympany PA, Puscinska E, Zielinski J, Welsh KI, du Bois RM. Analysis of MHC encoded antigen-processing genes TAP1 and TAP2 polymorphisms in sarcoidosis. *Am J Respir Crit Care Med* 1999; 160:1009-14.
76. Iannuzzi MC, Maliarik MJ, Poisson LM, Rybicki BA. Sarcoidosis susceptibility and resistance HLA-DQB1 alleles in African Americans. *Am J Respir Crit Care Med* 2003; 167:1225-31.
77. Sato H, Grutters JC, Pantelidis P, et al. HLA-DQB1*0201: a marker for good prognosis in British and Dutch patients with sarcoidosis. *Am J Respir Cell Mol Biol* 2002; 27:406-12.
78. Richeldi L, Sorrentino R, Saltini C. HLA-DPB1 glutamate 69: a genetic marker of beryllium disease. *Science* 1993; 262:242-4.
79. Maliarik MJ, Chen KM, Major ML, et al. Analysis of HLA-DPB1 polymorphisms in African-Americans with sarcoidosis. *Am J Respir Crit Care Med* 1998; 158:111-4.
80. Mungall AJ, Palmer SA, Sims SK, et al. The DNA sequence and analysis of human chromosome 6. *Nature* 2003; 425:805-11.
81. Hampe J, Shaw SH, Saiz R, et al. Linkage of inflammatory bowel disease to human chromosome 6p. *Am J Hum Genet* 1999; 65:1647-55.
82. Jenisch S, Henseler T, Nair RP, et al. Linkage analysis of human leukocyte antigen (HLA) markers in familial psoriasis: strong disequilibrium effects provide evidence for a major determinant in the HLA-B/-C region. *Am J Hum Genet* 1998; 63:191-9.
83. Gaffney PM, Kearns GM, Shark KB, et al. A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. *Proc Natl Acad Sci U S A* 1998; 95:14875-9.
84. Todd JA. Genetic control of autoimmunity in type 1 diabetes. *Immunol Today* 1990; 11:122-9.
85. Okamoto K, Makino S, Yoshikawa Y, et al. Identification of I kappa B as the second major histocompatibility complex-linked susceptibility locus for rheumatoid arthritis. *Am J Hum Genet* 2003; 72:303-12.

86. Haines JL, Ter-Minassian M, Bazyk A, et al. A complete genomic screen for multiple sclerosis underscores a role for the major histocompatibility complex. The Multiple Sclerosis Genetics Group. *Nat Genet* 1996; 13:469-71.
87. Amor M, Parker KL, Globerman H, New MI, White PC. Mutation in the CYP21B gene (Ile-172----Asn) causes steroid 21-hydroxylase deficiency. *Proc Natl Acad Sci U S A* 1988; 85:1600-4.
88. Dupont B, Oberfield SE, Smithwick EM, Lee TD, Levine LS. Close genetic linkage between HLA and congenital adrenal hyperplasia (21-hydroxylase deficiency). *Lancet* 1977; 2:1309-12.
89. Feder JN, Gnirke A, Thomas W, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* 1996; 13:399-408.
90. Ozaki K, Ohnishi Y, Iida A, et al. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 2002; 32:650-4.
91. Furuya K, Yamaguchi E, Itoh A, et al. Deletion polymorphism in the angiotensin I converting enzyme (ACE) gene as a genetic risk factor for sarcoidosis. *Thorax* 1996; 51:777-80.
92. Maliarik MJ, Rybicki BA, Malvitz E, et al. Angiotensin-converting enzyme gene polymorphism and risk of sarcoidosis. *Am J Respir Crit Care Med* 1998; 158:1566-70.
93. Papadopoulos KI, Melander O, Orho-Melander M, Groop LC, Carlsson M, Hallengren B. Angiotensin converting enzyme (ACE) gene polymorphism in sarcoidosis in relation to associated autoimmune diseases. *J Intern Med* 2000; 247:71-7.
94. Schurmann M, Reichel P, Muller-Myhsok B, et al. Angiotensin-converting enzyme (ACE) gene polymorphisms and familial occurrence of sarcoidosis. *J Intern Med* 2001; 249:77-83.
95. McGrath DS, Foley PJ, Petrek M, et al. Ace gene I/D polymorphism and sarcoidosis pulmonary disease severity. *Am J Respir Crit Care Med* 2001; 164:197-201.
96. Ziegenhagen MW, Muller-Quernheim J. The cytokine network in sarcoidosis and its clinical relevance. *J Intern Med* 2003; 253:18-30.
97. Zheng L, Teschler H, Guzman J, Hubner K, Striz I, Costabel U. Alveolar macrophage TNF-alpha release and BAL cell phenotypes in sarcoidosis. *Am J Respir Crit Care Med* 1995; 152:1061-6.
98. Seitzer U, Swider C, Stuber F, et al. Tumour necrosis factor alpha promoter gene polymorphism in sarcoidosis. *Cytokine* 1997; 9:787-90.

99. Higuchi T, Seki N, Kamizono S, et al. Polymorphism of the 5'-flanking region of the human tumor necrosis factor (TNF)-alpha gene in Japanese. *Tissue Antigens* 1998; 51:605-12.
100. Somoskovi A, Zissel G, Seitzer U, Gerdes J, Schlaak M, Muller Quernheim J. Polymorphisms at position -308 in the promoter region of the TNF-alpha and in the first intron of the TNF-beta genes and spontaneous and lipopolysaccharide-induced TNF-alpha release in sarcoidosis. *Cytokine* 1999; 11:882-7.
101. Hizawa N, Yamaguchi E, Furuya K, Jinushi E, Ito A, Kawakami Y. The role of the C-C chemokine receptor 2 gene polymorphism V64I (CCR2-64I) in sarcoidosis in a Japanese population. *Am J Respir Crit Care Med* 1999; 159:2021-3.
102. Petrek M, Drabek J, Kolek V, et al. CC chemokine receptor gene polymorphisms in Czech patients with pulmonary sarcoidosis. *Am J Respir Crit Care Med* 2000; 162:1000-3.
103. Spagnolo P, Renzoni EA, Wells AU, et al. C-C chemokine receptor 2 and sarcoidosis: association with Lofgren's syndrome. *Am J Respir Crit Care Med* 2003; 168:1162-6.
104. Tromp G, Kuivaniemi H, Raphael S, et al. Genetic linkage of familial granulomatous inflammatory arthritis, skin rash, and uveitis to chromosome 16. *Am J Hum Genet* 1996; 59:1097-107.
105. Hugot JP, Laurent-Puig P, Gower-Rousseau C, et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 1996; 379:821-3.
106. Hampe J, Schreiber S, Shaw SH, et al. A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am J Hum Genet* 1999; 64:808-16.
107. Cavanaugh J. International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16. *Am J Hum Genet* 2001; 68:1165-71.
108. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; 411:599-603.
109. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001; 411:603-6.
110. Hampe J, Cuthbert A, Croucher PJ, et al. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001; 357:1925-8.

111. Schreiber S, Nikolaus S, Hampe J. Activation of nuclear factor kappa B inflammatory bowel disease. *Gut* 1998; 42:477-84.
112. Drent M, van den Berg R, Haenen GR, van den Berg H, Wouters EF, Bast A. NF-kappaB activation in sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2001; 18:50-6.
113. Beutler B. Autoimmunity and apoptosis: the Crohn's connection. *Immunity* 2001; 15:5-14.
114. Costabel U, Hunninghake GW. ATS/ERS/WASOG statement on sarcoidosis. Sarcoidosis Statement Committee. American Thoracic Society. European Respiratory Society. World Association for Sarcoidosis and Other Granulomatous Disorders. *Eur Respir J* 1999; 14:735-7.
115. Livak KJ. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal* 1999; 14:143-9.
116. Kutuyavin IV, Afonina IA, Mills A, et al. 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res* 2000; 28:655-61.
117. Lantermann A, Gong J, Hampe J, Schreiber S. Development of a Single Nucleotide Polymorphism Map in the HLA-Region on Chromosome 6, German Human Genome Project (DHGP), Jahrestagung, Braunschweig, Germany, November, 2001.
118. Hampe J, Grebe J, Nikolaus S, et al. Association of NOD2 (CARD 15) genotype with clinical course of Crohn's disease: a cohort study. *Lancet* 2002; 359:1661-5.
119. Lesage S, Zouali H, Cezard JP, et al. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 2002; 70:845-57.
120. Zhang Y, Costin C, Giang C, Vorhaben R, Stastny P. Typing for all known HLA-DR alleles by group-specific PCR and flow cytometry-based multiplexed single nucleotide extension. *Hum Immunol* 2003; 64:S38.
121. Cordovado SK, Simone AE, Mueller PW. High-resolution sequence-based typing strategy for HLA-DQA1 using SSP-PCR and subsequent genotyping analysis with novel spreadsheet program. *Tissue Antigens* 2001; 58:308-14.
122. Geranton S, Rostagnat-Stefanutti A, Bendelac N, et al. High-risk genotype for type 1 diabetes: a new simple microtiter plate-based ELOSA assay. *Genet Test* 2003; 7:7-12.
123. Bodmer JG, Marsh SG, Albert ED, et al. Nomenclature for factors of the HLA system, 1990. *Hum Immunol* 1991; 31:186-94.

124. Marsh SG, Albert ED, Bodmer WF, et al. Nomenclature for factors of the HLA system, 2002. *Eur J Immunogenet* 2002; 29:463-515.
125. Moran P, Dightman DA, Park LK. Nonelectrophoretic genotyping using allele-specific PCR and a dsDNA-specific dye. *Biotechniques* 1998; 24:206-8, 210, 212.
126. Nurmi J, Kiviniemi M, Kujanpaa M, Sjoroos M, Ilonen J, Lovgren T. High-throughput genetic analysis using time-resolved fluorometry and closed-tube detection. *Anal Biochem* 2001; 299:211-7.
127. Wu J, Griffith BB, Bassinger S, et al. Strategies for Unambiguous Detection of Allelic Heterozygosity via Direct DNA Sequencing of PCR Products: Application to the HLA DRB1 Locus. *Mol Diagn* 1996; 1:89-98.
128. Wu J, Williams TM. DRB1/3/5 Locus Sequence Based Allele Identification. In: Tilanus MGJ, Hansen JA, Hurley C, eds. *IHWG Technical Manual Genomic Analysis of the Human MHC - DNA-Based Typing for HLA Alleles and Linked Polymorphisms*. International Histocompatibility Working Group, 2002.
129. Kimura A, Sasazuki T. Eleventh International Histocompatibility Workshop reference protocol for the HLA DNA -typing technique, 1991. Oxford University Press, Oxford.
130. Hampe J, Wollstein A, Lu T, et al. An integrated system for high throughput TaqMan based SNP genotyping. *Bioinformatics* 2001; 17:654-5.
131. Matisse TC, Sachidanandam R, Clark AG, et al. A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set. *Am J Hum Genet* 2003; 73:271-84.
132. Miceli-Richard C, Lesage S, Rybojad M, et al. CARD15 mutations in Blau syndrome. *Nat Genet* 2001; 29:19-20.
133. Wang X, Kuivaniemi H, Bonavita G, et al. CARD15 mutations in familial granulomatosis syndromes: a study of the original Blau syndrome kindred and other families with large-vessel arteritis and cranial neuropathy. *Arthritis Rheum* 2002; 46:3041-5.
134. Hardy G. Mendelian proportions in a mixed population. *Science* 1908; 28:49-50.
135. Weinberg W. On the demonstration of heredity in man. *Naturkunde in Wurttemberg, Stuttgart* 1908; 64:368-382.
136. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; 52:506-16.

137. Spielman RS, McGinnis RE, Ewens WJ. The transmission/disequilibrium test detects cosegregation and linkage. *Am J Hum Genet* 1994; 54:559-60.
138. Terwilliger JD. A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am J Hum Genet* 1995; 56:777-87.
139. Clayton D, Jones H. Transmission/disequilibrium tests for extended marker haplotypes. *Am J Hum Genet* 1999; 65:1161-9.
140. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; 58:1347-63.
141. Krawczak M, Konecki DS, Schmidtke J, et al. Allelic association of the cystic fibrosis locus and two DNA markers, XV2c and KM19, in 55 German families. *Hum Genet* 1988; 80:78-80.
142. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 1999; 22:139-44.
143. Rhodes DA, Stammers M, Malcherek G, Beck S, Trowsdale J. The cluster of BTN genes in the extended major histocompatibility complex. *Genomics* 2001; 71:351-62.
144. Stammers M, Rowen L, Rhodes D, Trowsdale J, Beck S. BTL-II: a polymorphic locus with homology to the butyrophilin gene family, located at the border of the major histocompatibility complex class II and class III regions in human and mouse. *Immunogenetics* 2000; 51:373-82.
145. Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet* 1992; 90:41-54.
146. Long M, Deutsch M. Association of intron phases with conservation at splice site sequences and evolution of spliceosomal introns. *Mol Biol Evol* 1999; 16:1528-34.
147. Sharpe AH, Freeman GJ. The B7-CD28 superfamily. *Nat Rev Immunol* 2002; 2:116-26.
148. Nurieva RI, Mai XM, Forbush K, Bevan MJ, Dong C. B7h is required for T cell activation, differentiation, and effector function. *Proc Natl Acad Sci U S A* 2003; 100:14163-8.
149. Ermann J, Fathman CG. Costimulatory signals controlling regulatory T cells. *Proc Natl Acad Sci U S A* 2003; 100:15292-3.
150. Dong C, Nurieva RI, Prasad DV. Immune regulation by novel costimulatory molecules. *Immunol Res* 2003; 28:39-48.

151. Hampe J, Wienker T, Nurnberg P, Schreiber S. Mapping genes for polygenic disorders: considerations for study design in the complex trait of inflammatory bowel disease. *Hum Hered* 2000; 50:91-101.
152. Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl* 1995; 4:357-62.
153. Yan SD, Chen X, Fu J, et al. RAGE and amyloid-beta peptide neurotoxicity in Alzheimer's disease. *Nature* 1996; 382:685-91.
154. Hofmann MA, Drury S, Fu C, et al. RAGE mediates a novel proinflammatory axis: a central cell surface receptor for S100/calgranulin polypeptides. *Cell* 1999; 97:889-901.
155. Wei J, Hemmings GP. The NOTCH4 locus is associated with susceptibility to schizophrenia. *Nat Genet* 2000; 25:376-7.
156. McGinnis RE, Fox H, Yates P, et al. Failure to confirm NOTCH4 association with schizophrenia in a large population-based sample from Scotland. *Nat Genet* 2001; 28:128-9.
157. Henry J, Miller MM, Pontarotti P. Structure and evolution of the extended B7 family. *Immunol Today* 1999; 20:285-8.
158. Bernard CC, Johns TG, Slavin A, et al. Myelin oligodendrocyte glycoprotein: a novel candidate autoantigen in multiple sclerosis. *J Mol Med* 1997; 75:77-88.
159. Breithaupt C, Schubart A, Zander H, et al. Structural insights into the antigenicity of myelin oligodendrocyte glycoprotein. *Proc Natl Acad Sci U S A* 2003; 100:9446-51.
160. Borriello F, Sethna MP, Boyd SD, et al. B7-1 and B7-2 have overlapping, critical roles in immunoglobulin class switching and germinal center formation. *Immunity* 1997; 6:303-13.
161. Collins AV, Brodie DW, Gilbert RJ, et al. The interaction properties of costimulatory molecules revisited. *Immunity* 2002; 17:201-10.
162. Walunas TL, Bakker CY, Bluestone JA. CTLA-4 ligation blocks CD28-dependent T cell activation. *J Exp Med* 1996; 183:2541-50.
163. Carreno BM, Collins M. The B7 family of ligands and its receptors: new pathways for costimulation and inhibition of immune responses. *Annu Rev Immunol* 2002; 20:29-53.
164. Agadjanyan MG, Chattergoon MA, Holterman MJ, et al. Costimulatory molecule immune enhancement in a plasmid vaccine model is regulated in part through the Ig constant-like domain of CD80/86. *J Immunol* 2003; 171:4311-9.

165. Zissel G, Ernst M, Rabe K, et al. Human alveolar epithelial cells type II are capable of regulating T-cell activity. *J Investig Med* 2000; 48:66-75.
166. Ueda H, Howson JM, Esposito L, et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 2003; 423:506-11.
167. Luisetti M, Beretta A, Casali L. Genetic aspects in sarcoidosis. *Eur Respir J* 2000; 16:768-80.
168. Kirsten D. Sarcoidosis in Germany. Analysis of a questionnaire survey in 1992 of patients of the German Sarcoidosis Group. *Pneumologie* 1995; 49:378-82.
169. McGrath DS, Daniil Z, Foley P, et al. Epidemiology of familial sarcoidosis in the UK. *Thorax* 2000; 55:751-4.
170. Rybicki BA, Maliarik MJ, Bock CH, et al. The Blau syndrome gene is not a major risk factor for sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* 1999; 16:203-8.
171. Cuthbert AP, Fisher SA, Mirza MM, et al. The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. *Gastroenterology* 2002; 122:867-74.
172. Consolandi C, Busti E, Pera C, et al. Detection of HLA polymorphisms by ligase detection reaction and a universal array format: a pilot study for low resolution genotyping. *Hum Immunol* 2003; 64:168-78.
173. Feolo M, Fuller TC, Taylor M, Zone JJ, Neuhausen SL. A strategy for high throughput HLA-DQ typing. *J Immunol Methods* 2001; 258:65-71.
174. Slateva K, Elsner HA, Albis-Camps M, Blasczyk R. HLA-DRB fluorotyping by dark quenching and automated analysis. *Tissue Antigens* 2001; 58:250-4.
175. Hunninghake GW, Gadek JE, Kawanami O, Ferrans VJ, Crystal RG. Inflammatory and immune processes in the human lung in health and disease: evaluation by bronchoalveolar lavage. *Am J Pathol* 1979; 97:149-206.
176. Roitt I, Brostoff J, Male D. *Immunology*; sixth edition. 2001;87-103.
177. Alberts B, Jochanson A, Lewis J, et al. *Molecular biology of the cell*; fourth edition. 2001; 1392-1409.

9. Appendix

9.1. Supplemental data

9.1.1. Primer and probe information for chromosome 6 assays

Table 15. Part 1: Primers and probes of SNP assays

Marker	hCV number*	Gene	Position (NCBI) MB	Primers: Forward 5'-3' Reverse 5'-3'	Probes	Sample used
rs2073528	hCV11255445	<i>BTN3A2</i>	26,44			Exploratory/extension
rs2072803	hCV2474923	<i>BTN2A2</i>	26,46			Exploratory/extension
-	hCV2474907	<i>BTN3A1</i>	26,48			Exploratory/extension
-	hCV2474883		26,50			Exploratory/extension
-	hCV9804528		26,54			Exploratory/extension
rs3736781	hCV9804725	<i>BTN1A1</i>	26,57			Exploratory/extension
rs126007			27,80	TCCTTAACCAACTCTACTCTGG TGGTAACCTGACAGCTGATCTTGT	FAM-AGAAGAGCAGCAACCAAGGCTAACGTGA TET-AGAAGAGCAGCAAAACAAGGCTAACGTGAGA	Exploratory-screen
rs929042		<i>RFP</i>	28,95	CGGACGTGTGGTTCACAGC CTCCGTGTACAAATGTGTCCGTG	FAM-CTGTAGCTGGCGAGGAAGCCAGGAGT TET-CTGTAGCTGGCGAGGAAGCCAGGAGT	Exploratory-screen
rs53161			29,56	GGGCTGTCCACGTGCAC GCAGAGTTGAACAGTATATAACCTTTC	FAM-AAGGGGCAACATACACAATGTGTTACCAAAATT TET-AAGGGGCAACATACACAATGTGTTACCAAAATT	Exploratory-screen
rs404240		<i>UBD</i>	29,59	CCACAAGAAAACAAGGCGAGC GAACATGTCCGGTCTAAGACCAA	FAM-TCTTGGAGCCCGAGCAAGAAAGACCTG TET-ATCTTGGAGCCCGAGCAAGAAAGACCTGG	Exploratory-screen
rs362536		<i>UBD</i>	29,59	TGGAGAGAACCCTGGAATTAAGGC GGACTATATCTAAGTCTTACTTCAATAGCTGGA	FAM-ATTGGGAGAATATTCTGAATGGAGTATGAGGGAG TET-TATCTAGATTTGGGAAATATTCTGAATGGAGTATGAGGG	Exploratory-screen
rs1805057		<i>GABBR1</i>	29,64	CCATCTGGGCTCTGGCA GGTAATGGTCTGGTGTGTGTAGTTGAA	FAM-AGGAGGCGGCGCTCTGGTGTG TET-TGGAGGCGGCGCTCTGGTGTG	Exploratory-screen
rs3130253		<i>MOG</i>	29,73	GTTCTCCTCGGGTGTGCTG CCCACCCTCGCCCTGTAC	FAM-CACTCTTGGCCTCGTCTTCTCTGCGC TET-TCACTCTTGGCCTCATCTTCTCTGCGCT	Exploratory-screen
rs1569315		<i>HLA-G</i>	29,85	CGCAGGTCCTCGTTCAGG AGTGGATGATTTGGCTCGCA	FAM-AGGAGGCTCCCTCGGACCC TET-AGGAGGCGCGCTCGGACCC	Exploratory-screen
rs8347		<i>HCG-V</i>	30,10	AGGAGACCTCAAGCTCCAG AAGTGACCAATCTCAAGCCAGC	FAM-AAATGGAAATAGAATAATCATATCTTCTTCCACCC TET-AAATGGAAACAAGAATAATCATATCTTCTTCCACCC	Exploratory-screen
rs1264457		<i>HLA-E</i>	30,52	CTCACACCCTCGAGTGGATG AGGGTGAGATATCTTCCCG	FAM-CCCAGCAGGCGCTTCTCCCG TET-CCCAGCAGGCGCTTCTCCCG	Exploratory-screen
rs2074510		<i>GTF2H4</i>	30,94	GAGTGCAGCGTCTATCTCATCC TTGGCTCTGGAGATGAGG	FAM-CATCTTGGGCTCTTGCCTGTGAAATCTTTC TET-CCATCTGGGCTCTTCTCTGAAATCTTTC	Exploratory-screen/extension
rs3095318		<i>CDSN</i>	31,15	CGTCTGGGCAACCTG AGACACACAGCTAGAGGTAGGTGCTCA	FAM-CACGGGATGATGGCACTGCTGCT TET-CACGGGATGATGGCACTGCTGCT	Exploratory-screen
rs2073721		<i>SC1</i>	31,19	GGTGGACCAAGAGCTGCG CATCCAGTTCCGCCAACAC	FAM-ACCTGGGAAAGTGGGACACAG TET-ACCTGGGAAATGGGACACAG	Exploratory-screen
rs1050393		<i>HLA-C</i>	31,30	GGAGTATTGGGACCGGAG CGCTCTGGTGTAGTAGCCG	FAM-CCAGGCACAGCTGACCGAGT TET-CGCCAGGCACAGACTGACCGAGT	Exploratory-screen
rs1051488		<i>HLA-B</i>	31,38	CTGGCCTGGCTGTCTAGC CTTCCCAGTAATGAGGACGGG	FAM-AGCTGTGTGCTGCTGTGTGATGTA TET-AGCTGTGTGCTGCTGTGTGATGAGG	Exploratory-screen
rs1063630		<i>MICA</i>	31,44	CCCAGAGCCCCACAGTCTT AGAAACCCTGACTGCACAGATC	FAM-ATCCCCGACAGCA VIC-CATCCAGGACAGCA	Exploratory/extension
rs1063635		<i>MICA</i>	31,44	CTTTGAGCCACGACCCCA GGCTCACCCAGAGGGCACA	FAM-ACCAGGATTTGCCAAGGAGAGGAGCA VIC-ACCAGGATTTGCCAGGAGAGGAGC	Exploratory-screen/extension
rs3828916	hCV3273671		31,53			Exploratory/extension
rs3131639		<i>MICB</i>	31,53	GCAGAAACCGCAGGCAAA TGTCAAGTCTCTGGTCTCTGTGT	FAM-CAGGACATTTTCTG TET-AGGACATCTTCTGCC	Exploratory/extension
rs3134900		<i>MICB</i>	31,53	CTTGGGTGAAAGGTGATGG GGTTTGGGAGAGGAGAGCTC	FAM-TCTGTGAGATCCATGAAGACAGCAGCACC TET-CTGTGAGATGCATGAAGACAGCAGCACC	Exploratory-screen/extension
AC006046_G3 0156A		<i>MICB</i>	31,53	CTGGCTCTGCCCTTCTCTC CGGTGATGTGGCCCTCTGA	FAM-TGCAGATGACATTC TET-TGCAGGTGACATTC	Exploratory/extension
rs3130062		<i>IkBL</i>	31,59	CGGGAACATGCCAGGAGT GGCTGGAGCCCTCAGCA	FAM-TGCAGGATCCCTCTGCTTCTG TET-TGCAGGATCCCTCTGCTTCTG	Exploratory-screen/extension
rs1800683	hCV7514865	<i>LTA</i>	31,60			Exploratory/extension
rs1041981		<i>LTA</i>	31,60	CCGTCAGCACCCCAA GATG TGGGAGTCAAGTGGATGTTTACC	FAM-TGCCACAGCAACCTCAACCTGC TET-TTGGCCACAGCACCTCAACCTGC	Exploratory/extension
rs1799964		<i>TNPa</i>	31,60	GATGGGACTCACCAAGGTGAGG CCAGAGGTCTCTGTAAACCA	FAM-AGCAAAGGAGAGCTGAGAAGATGAAGGAAAGTC TET-AGCAAAGGAGAGCTGAGAAGATGAAGGAAAG	Exploratory-screen/extension

Table 15. Part 2: Primers and probes of SNP assays

Marker	hCV number*	Gene	Position (NCBI) MB	Primers: Forward 5'-3' Reverse 5'-3'	Probes	Sample used
rs1799742		TNPa	31,60	TCCAGGGCTATGAAAGTCGA TGCTGGTTTCAGTCTTGGCT	FAM-AGTATGGGGACCCCTTAATGAAGA TET-TATGGGGACCCCTTAACGAAG	Exploratory-screen
rs1800629		TNPa	31,60	CCTGCATCCTGTCTGAAGTTAGAAG TGGCCACTGACTGATTGIGTGT	FAM-AACCCCGTCCCTCATGCCCTC TET-AACCCCGTCCCTCATGCCCTC	Exploratory-screen
rs361525		TNPa	31,60	CAGTGGCCGAGAAGACCC AGCATCAAGGATACCCCTCAC	FAM-AATCAGAGCAGGAGGATGGGA TET-AATCGGAGCAGGAGGATGGG	Exploratory-screen
rs707916		DDAH2	31,76	CTGGAGCCGGGAGTATGTC CGAGCCTAAGGAGTTAAGCATCC	FAM-TTTATACTAGGTCTGTCTAGGGGACGCT TET-TTTATACTAGGTCTGTCTAGGGGACGCT	Exploratory-screen
rs707926		VARS2	31,81	GGTGGAGTGGGAGCAGTCAG TCACGGCCGGAAGATGAG	FAM-CCATAGATGACGTCCAGGGGATCGAT TET-CCATAGATGACATCCAGGGGATCGATG	Exploratory-screen
rs2075800		HSP70 - HOM	31,84	AACGAGCTCCTTTCGTGGCT CACATACCTCTGTCCGACAG	FAM-CATAAGAGAAGGAATTGGAGCAGATGTGTAACCC TET-TCATAAGAGAAGGAATTGAAGCAGATGTGTAACCC	Exploratory-screen
rs539689		HSPA1B	31,86	CAAAGTCTTGTAGTCCACAGT AACCCCATCATCAGCCGA	FAM-CCCCAAGCCGACAGGCC TET-CCCCAAGCCGACAGGCC	Exploratory-screen
AF019413_C8 9532G		C2	31,96	TGAGCGTGGCATTATCAC TCCTCTCTCATCACCTCAAGT	FAM-TCCCGGATATGACTGAGGTGATCAGC TET-CCCGGATATGACTGAGGTGATCAGC	Exploratory-screen
rs1035798		AGER	32,18	CCAGGGTCTTCTCCAGGGC GCOGTGAGTTCAGAGCAG	FAM-TGGCTCATTTCCACATACAGAGTTGAGAAC TET-TGGCTCATTTCCACACACAGAGTTGGA	Exploratory-screen
rs2070600		AGER	32,18	AGGCCCTGGGACAGTGT CCGACAGCCGGAAGGAA	FAM-CCAACGGCTCCCT TET-CCAACGGCTCCCTC	Exploratory/extension
rs1800624	hCV3293837	AGER	32,18			Exploratory/extension
rs2071287	hCV3293818	NOTCH4	32,20			Exploratory/extension
rs384247	hCV2412441	NOTCH4	32,21			Exploratory/extension
rs915894	hCV3293780	NOTCH4	32,22			Exploratory/extension
rs436845	hCV2412414		32,22			Exploratory/extension
rs560505	hCV2495657	C6orf10	32,29			Exploratory/extension
rs485774	hCV2495577	C6orf10	32,32			Exploratory/extension
rs552339	hCV2495574	C6orf10	32,32			Exploratory/extension
rs761187	hCV2234065	C6orf10	32,34			Exploratory/extension
rs2073044	hCV2488525	C6orf10	32,37			Exploratory/extension
rs2050190	hCV2488524	C6orf10	32,37			Exploratory/extension
rs2073045	hCV2488523	C6orf10	32,37			Exploratory/extension
AL034394_T2 6355G			32,37	AGAACTTCTTAGATGAGAGTGCACTTC GGGAAGAAAATGCTATGGCAAAAGG	FAM-TGAGCGTAGTCGATTC TET-TGAGCGTAGTCGATTC	Exploratory/extension
AL034394_A2 6270C			32,37	CCATAGACATTTTCTTCCCTGATGATTTTT CTTGAGACACAGATTTCTGGAACTT	FAM-CAACACATTTTCTTTTC TET-CAACACATTTTCTTTTC	Exploratory/extension
AL034394_C2 0283T			32,38	GTAGGAGCTTAAGACACTGTGACTG GCCAAGCCTCTGAAGACTAAGAGA	FAM-AGGGCTATTGCTTATT TET-CAGGCTATTACTTATT	Exploratory/extension
AL034394_C1 6965T			32,38	GCAGCTCATCCCAACTCATC GCAGCTCATCCCAACTCATC	FAM-TTCTTCAACGGGAGATG TET-TTCTTCAACAGAGAATG	Exploratory/extension
-	hCV2488476		32,38			Exploratory/extension
AL034394_C1 2339T			32,39	GTTCAGAGTGTCTTTCATGAGTAT CAGGATGGCAGAGCTCAGA	FAM-CGCATGGGACCGC TET-CGCATGGGACCGC	Exploratory/extension
AL034394_T1 0552C			32,39	ACTGTAAGAAAGAGAATCCATTTCTGATAATTA ATCAATAT GCTGTTTTGTGTCTCTAGAAGATG	FAM-TTCCACTTTAAATTTCTTACGTC TET-CTTCCACTTTAAATTTCTTATGTC	Exploratory/extension
BTNL2_E360A mb		BTNL2	32,39	GCGAGGAGAAAATCGCACTT TTTCCAAACCTGAAGGAACATAAGGAA	FAM-TTCTCTCTCATGTTGGTG TET-TTCTCTCTCATGTTGGTG	Exploratory/extension
BTNL2_P299Q		BTNL2	32,39	GAGTCAGGGCCCTGGGAAGA GGAGGGGCAAGAAGATGGA	FAM-TGTCCCTCCACTGCA TET-ATGTCCTCCATGGC	Exploratory/extension/replication
BTNL2_M286I		BTNL2	32,39	ACCCCTTAGCAATGTCTGCACGT GGTTCTTCCCACTGATCACTG	FAM-ATCCGAAGAGCATATC FAM-CATCTGAAGAGCATATC VIC-CATCTGAAGAGCACATC VIC-ATCCGAAGAGCACATC	Exploratory/extension/replication
BTNL2_P285L		BTNL2	32,39	GTCTCTCCCACTGATCACTGT TGAGTCAGGGCCTGGGA	FAM-TGGAGAAATGCAGCTGA TET-ATGGAGAAATGCAGCCGA	Exploratory/extension/replication
rs2076530		BTNL2	32,39	AGTACCCTGCTCTTTTGAATA GTGGCAGGAGCAGGTATTGAATA	FAM-AAGGTGGTAGTAAAGAA TET-AGGTGGTAGTAAAGAA	Exploratory/extension/replication
BTNL2_S240L		BTNL2	32,39	AGACTGACCCCTGAGATACTCA TTTTCAAAGGCGAGGTACTG	FAM-TCGTCCAAAGGTCT TET-TCGTCCAAAGGTCT	Exploratory/extension

Table 15. Part 3: Primers and probes of SNP assays

Marker	hCV number*	Gene	Position (NCBI) MB	Primers: Forward 5'-3' Reverse 5'-3'	Probes	Sample used
rs2076529		<i>BTNL2</i>	32,39	GGAGAGCAGATGGCAGAGTACAG CCTCGTCAATGGCGTCACT	FAM-ACCAGCACAGTCTCT TET-ACCAGTACAGTCTCTC	Exploratory/extension
rs2294878	hCV2488463	<i>BTNL2</i>	32,39			Exploratory/extension
<i>BTNL2</i> _i4_A2 39T		<i>BTNL2</i>	32,40	TTGCAGCTGTGTACCTGACA CAGAGATGATGCCATGCTTCT	FAM-ATCAGATGGTCACTTAT TET-CATCAGATGGACACTTAT	Exploratory/extension
rs2076524	hCV2488453	<i>BTNL2</i>	32,40			Exploratory/extension
<i>BTNL2</i> _i3_C1 900T		<i>BTNL2</i>	32,40	TCTGAGCATCGCATCCAAGAT AGACTCTGCGAGGGCGTTCCT	FAM-TGGCTTCCACATAGA TET-CTCCGCATAGAAC	Exploratory/extension
rs2076523		<i>BTNL2</i>	32,40	CCAGAGCCCCAGGTGTATTG TGGACCAAGCAGGACACAGA	FAM-CATCGCATCCAAGATAAAGATGGCTGT TET-CATCGCATCCAAGATGAAGATGGCCCT	Exploratory-screen/extension
<i>BTNL2</i> _i3_A1 837G		<i>BTNL2</i>	32,40	CAGAGCCCCAGGTGTATTG ACACGGCCAGCAGCTTCT	FAM-CTCCCCGATGTCT TET-CTCCCTGGATGTCT	Exploratory/extension
<i>BTNL2</i> _W94R		<i>BTNL2</i>	32,40	GACTGAGTGCAGATGGAGGATAC CTTTGCAATGCCATTTCTATCC	FAM-AGAGGCTGGGTAGAGT TET-AGAGGACGGGTAGAGT	Exploratory/extension
<i>BTNL2</i> _H60H		<i>BTNL2</i>	32,40	TGCCAGTACTCCCAAGAG GGGCTCTGAGCGGTACCA	FAM-AATGCATGTGGAGGTG TET-AATGCACGTGGAGGT	Exploratory/extension
<i>BTNL2</i> _i1_G1 717T		<i>BTNL2</i>	32,40	GGAGCCAAAGCCCAATTTTCC CAGTCTTAGCAGAGATCCACATCA	FAM-ACCCTGAACAGTCT TET-CCCTGCACAGTCT	Exploratory/extension
	hCV2455668		32,43			Exploratory/extension
DRAp_231		<i>HLA-DRA</i>	32,43	AATGTGCTTCAGGTATATCCCTGTCT AGTAAAGTCTTTAAACAACAGGACAAACAA	FAM-TGTCCGTGATTGAC TET-TGTCCGTCAATGAC	Exploratory/extension
DRAp_224		<i>HLA-DRA</i>	32,43	AATGTGCTTCAGGTATATCCCTGTCT AGTAAAGTCTTTAAACAACAGGACAAACAA	FAM-ATTTAAGACTGTTAGTCAAT TET-TTAAGACTGTTAGTCAAT	Exploratory/extension
DRAp_196		<i>HLA-DRA</i>	32,43	AATGTGCTTCAGGTATATCCCTGTCT CCGTTTCATTTGGATAAAGAAGTAAAGT	FAM-CAACAACAACAACAA TET-ACAACGACAAACAAT	Exploratory/extension
DRAp_127		<i>HLA-DRA</i>	32,43	TTATCCAATGACCGGATATCTTTGTG ATGAGCATCTGTGTCTAGGG	FAM-CCTGGACTCTTTTG TET-CTGGACCCCTTTGCA	Exploratory/extension
rs14004		<i>HLA-DRA</i>	32,43	CCTCACTCCCGAGCTCTACTGA CCACTTATGGCCATTTCTCTCTG	FAM-CTCCCAACAGAGCG TET-TOCCAAAGAGCGC	Exploratory/extension
	hCV2455646	<i>HLA-DRA</i>	32,44			Exploratory/extension
rs8084	hCV2455637	<i>HLA-DRA</i>	32,44			Exploratory/extension
rs7192		<i>HLA-DRA</i>	32,44	TGGGCTGACTGTGGGTC ACCACACTTAACCTCACTCCATG	FAM-TCATCATCAAGGGATTGGCCAAAAGC TET-TCATCATCAAGGGATTGGCCAAAAGC	Exploratory-screen/extension
rs2213584		<i>HLA-DRA</i>	32,44	AATGGGTAAGGCCAGTCTTCA TGCAGGATTTACATATCAACATCGT	FAM-TGTCCAATCTCTTTG TET-TGTCCAACCTCTTTG	Exploratory/extension
rs2395182		<i>HLA-DRA</i>	32,44	GGTGGTTTCAAGAATCAGTCAGACT GGCCTTACCCATTTCTGTAGACATA	FAM-CGGTAATCAATAGGC TET-CTCGTAAATAAATAGG	Exploratory/extension
rs707952		<i>HLA-DQA1</i>	32,62	CATGAGGTTCTGAGGTCACA TGGTCTCAGAAACACCTCTCTGT	FAM-CTGGGTCAGCCCAACACCTCT TET-CTGGGTCAGCCCAACACCTCTCA	Exploratory-screen/extension
X66401_G486 29A		<i>TAP2</i>	32,80	GGTGGTTTCAAGTGTGGGAG CTCCATTTCTGGATGAAGTCA	FAM-TGTGGGAACAACATTACTTATGGGCTGAG TET-TGGGAACAACATTGCTTATGGGCTGC	Exploratory-screen
rs1042337		<i>HLA-DMB</i>	32,91	GCCCAGCCCAATGGAGAC CCTAATTCATCCATCTGCCATAC	FAM-TGTGTGGTAGACACACTGGGGCTCC TET-TGTGTGGTAGACACACTGGGGCTCC	Exploratory-screen
rs1063478		<i>HLA-DMA</i>	32,92	AAAAGGCTGGAAGCTGAGTC GAGTTTGGCAAGCCCAACAC	FAM-CTTCCACAGGGATGGAATGATGCTG TET-CTTCCACAGGGACGGAATGATGCT	Exploratory-screen
rs3918143		<i>RING3</i>	32,95	GATGCCAGATGAACACTAGAAC CCTCAGAGAGCTCTCACTGCTAC	FAM-CCTGGCTTGGCCAAATGCTCTCAG TET-CCTGGCTTGGTCAAATGCTCTCAGA	Exploratory-screen
rs2581		<i>HLA-DNA</i>	32,98	TGTGCTGACTTTGAGTGGGATC AACACCACCAAGCATTTAGTGC	FAM-CCTTCCCAATCAACACACACACATTC TET-CCTTCCCAATCAACACACACACATTTCT	Exploratory-screen
rs926424		<i>RXRβ</i>	33,16	CTCCTCTCTTCCACTGATGTGC GCACGTGAAGACTGTGGC	FAM-CATTGCTCTCTTTCACACCGATCC TET-CATTGCTCTCTTTCACACCGATCC	Exploratory-screen
rs1547387		<i>HKE4</i>	33,17	AGTGGGGAGGTGAGGAATAG CCCCAATACACACTCATTTGTGC	FAM-CCTCACCCTGTGGGAGTGTCCATGT TET-CTCACCCTGTGGGAGTGTCCATGT	Exploratory-screen
rs466384		<i>BING4</i>	33,26	CCTCAGGCTTAGGTGTATTAGCA TCTGCTGTAGACTACTGCCC	FAM-CTCTCCAGTACTGTGTCTTATGGAGTCCAG TET-TCCAGTACTGCTCTTATGGAGTCCA	Exploratory-screen
rs2071888		<i>TAPBP</i>	33,27	GCATTTGCTGTCTGGGATG AGCAAGCTCCAGGGTGAAC	FAM-CTTCTGGCTGCTAGAGTTCACCCCTTTCA TET-TCTGGCTGCTCAGTTCACCCCTTTCA	Exploratory-screen
rs999942		<i>ITPR3</i>	33,63	CCCTCCATCAGACTGGCCTC CGGTCAACAGCCACATGC	FAM-CAGCAGGAGGCTTGCATGG TET-CAGCAGGAGGCTTGCATGG	Exploratory-screen
rs943461		<i>MLN</i>	33,77	CCCCAAGCCAGAGAGACC AGGATACAGCTCATGCTGAA	FAM-CAAGCCAGATGACGAGGGCTCTG TET-CAAGCCAGATGACGAGGGCTCTG	Exploratory-screen
rs755658		<i>FKBP5</i>	35,55	TTGAAGTCTGTATCACAGGCATG ACCAACAGTGTAAAGCCMTGTTACTAGG	FAM-ACTCGACATGAGATGTACGCGCTGC TET-TGGCAGATGAGATGTACGCGCTGC	Exploratory-screen
rs992105		<i>FKBP5</i>	35,56	TTCCATCATGACCAGCTGTAGC CAGAGGAGATAAGGCATGTC	FAM-ACTCTGTGTAAGCAGCTACAAAGTGCATCTTTAGAA TET-CGTACTCTGTGTAAGCAGCTACAAAGTGCATCTTTAGAA	Exploratory-screen

Table 15. Part 4: Primers and probes of SNP assays

Marker	hCV number*	Gene	Position (NCBI) MB	Primers: Forward 5'-3' Reverse 5'-3'	Probes	Sample used
rs1320991		FKBP5	35,56	TGTTGGTATGTTGGAGAATTATGTGAA AAACAAGTAAATTCACCTGCCTGCAC	FAM-TAGAAAACCCCTACGTGCTATATATGTGTGCGT TET-TAGAAAACCCCTACATGCTATATATGTGTGCGTATG	Exploratory-screen
rs1326752		SRPK1	35,83	AAAAAGCCTGTTGTCTTAGGTATGC AGAGGTTAATGCTAGGTTGCGC	FAM-CTGCAGCCAGGTCAGCCCGAA TET-CTGCAGCCAGGTCAGCCCGA	Exploratory-screen
Z95152_G577 23A			36,03	CAAGTGAATGGCTGTTGACG CCCAAAAAGAGGTCATGATAGG	FAM-TGAGTAAATTTTTCACATTTGCTCTCTGGTCTA TET-TGAGTAAATTTTTCGCTTTGCTCTCTGG	Exploratory-screen
Z95152_T587 84G			36,08	GGAGAAGAGGGCTAATATATCCTAGAGTAG GGCCGCTGTAATCTCTTACTCTG	FAM-TTCTCTCTCTCTCTGTTACTGATAGCTGCTGACTT TET-TTCTCTCTCTCTCTGTTACTGATAGCTGCTGAC	Exploratory-screen
Z95152_T106 861C			36,08	GAAGTCATCAGCTTTGTGCCAC ACCATGGAGGAATCTCTGCA	FAM-AAGGCCTTTTCACGGGAACCTCTCCAAA TET-AAGGCCTTTTCATGGGAACCTCTCCAAAATTC	Exploratory-screen
rs763096		MAPK13	36,10	CTGTGAGGACTGGCAGATGTG GCCTTTGCTGCTCAAGTGTG	FAM-CCAGGCAAGCACTCCAAAAGTCTT TET-CAGGCAAGCACTCCAAAAGTCTT	Exploratory-screen
rs2007683		MAPK13	36,10	CAGCTTCTTCCATCTCTCA ACAGTCGATATGGGTAGCCA	FAM-ATGTTGATATCTGAATCAATGCTATCTTACCACATT TET-TGTTGATATCTGAATCAATGCTATCTTACCACATTCTACA	Exploratory-screen
rs881648		TEL2	36,35	CTAGGCCTCAACCAGGAC CCCATGGCTTATTACAGCTAGGA	FAM-CTCAGTTTGGATGACAGGAAATGAGACTTC TET-CTCAGTTTGGATGACAAAGGAAATGAGACTTC	Exploratory-screen
rs881647		TEL2	36,35	TTGAGAACTGGCTGATATCAGC AATGAGGCATGGTGTGAC	FAM-CAAGGACAACCTGGGAGAATTTGAGAGG TET-AAGGCACACTGGGAGAATTTGAGAGGA	Exploratory-screen
rs916287		TEL2	36,35	AGAAGAAGAGACGACAGGACCG CCATTTTCAGAACTGGGATGATA	FAM-CTCTGAACTGCTCTCTCCTCAAAAATCTCCCAA TET-CTCTGAACTGCTCTCTCCTCAGAAATCTCCCAA	Exploratory-screen
rs1885205		TEL2	36,35	CTCCTGACTCTACTCTGGCTTG TCCACTCTGAGGACAACCGTG	FAM-ACCCTCATGGCCATGGACCGG TET-CATGGCCATGGACAGGATGG	Exploratory-screen
rs2055272		PLA2G3B3	37,29	CATGGTGGTAATGAATTTGTGATTATG ATGCTTAAGATTTATGTCAATGATCCCA	FAM-AGGTAGGTTTCGGGCTCTTACATTTGCT TET-AGGTAGGTTTCGGGCTCTTACATTTGCTCCAC	Exploratory-screen
rs1004279		GLP1R	39,03	AGGAGGCTGAGGTCAGGC GCCATCATGAACTTATCACTACTGAC	FAM-TGAGTGTCCGCTGCTCTTCC TET-AGTGTGTCCGCTGCTCTTCC	Exploratory-screen
rs1541816		KCNK5	39,16	AGGCCTCTAGAAATGGACCA GAGCAGGAGGCCAAGTCAGA	FAM-CCCTCCCGGTTAGGCTGGGTAC TET-CCCTCCCGGTCAGGCTGGG	Exploratory-screen
rs1014737		APOBEC2	41,02	AACACAGAGGGAGGTAGGTCA GTGTGATTATAAAGCCCTGGACA	FAM-AGCAATGGATCCTGAAGGGTTAAGTCCACT TET-AGCAATGGATCCTGAAGGGTTAAGTCCACT	Exploratory-screen
rs2234246		TREM1	41,25	AAAAGGCAGGGAGTTAATAACATGA TAGAGGACGAGGGCAGTGG	FAM-ATCTGTAATCAGCGGCTATTTCTAAGTCAAGCTCT TET-ATCTGTAATCAGCGGCTATTTCTAAGTCAAGCTCTCA	Exploratory-screen
rs1817538		TREM1	41,25	CTGCTTGCATGTCTGTGATG CGAGAAGAATCCACCGTAGTG	FAM-CCAGAAAACCTTAGAATCATTTTGCCAAAATCT TET-CCAGAAAACCTTAGAATCATTTTGCCAAAATCT	Exploratory-screen
rs1351835		TREM1	41,25	ATGTTCTGATGATGACTGGC AGGAGTGAAGGCACTTGGC	FAM-CTCTCATCTGCTTGTGTGTGCTGATGAA TET-TCTCATCTGCTTGCATGTCTGCTGATGA	Exploratory-screen
rs2234237		TREM1	41,25	GAACCCATTCTCTTTCCCTGCT AGCGTGTAGTCACTTTCACATCC	FAM-CTCCGAGCTGCAACTAAATTAATCTGAGGAAAAGTATGA TET-CTCCGAGCTGCAACTAAATTAATCTGAGGAAAAGTATGA	Exploratory-screen
AL391903_C8 8694T		TREM1	41,26	ATTACATTCAGATATCATTTCCCTTAGAT GACATCAACACAATAATACAGAGACCAA	FAM-AGGTAATTTGCTATTTACCACAAAAGGAAAAGTGGAG TET-AGGTAATTTGCTATTTACCACAAAAGGAAAAGTGGGA	Exploratory-screen
rs742735		MDPI	41,62	TCAATTGATGGCTGATGAGGAG AGCAGAATAAAATGGTCCAGG	FAM-CACATCGGCTGGTGGGTGCTG TET-CACATCGGCTGGTGGGTGCTG	Exploratory-screen
rs1410492		CCND3	41,91	GGCATTGTGGTGTGA TGG GCCAAAGTACTAGTGTAGACATGCA	FAM-AGGGGATGCCACTCTCTGCTGGTCT TET-AGGGGATGCCACTCTCTGCTGGTCTT	Exploratory-screen
rs1160867		TNRC5	42,90	CCCTAAGCAGGAGCTCTGG TTCCCCACATCTTCCAGCAC	FAM-ACCAAGACCCGAGGAGAAGAGCT TET-CAAGACCCGAGGAGAGGCTGGG	Exploratory-screen
rs9395		SRF	43,15	CCCAGTGAATGAATCAGGCA ACCCAGAGCCACCCAG	FAM-ATGCGTGTGTCAGGATGAGTTGAGGT TET-CGTGTGTGTCAGGATGAGTTGAGGTGA	Exploratory-screen
rs1418488		CAPN11	44,14	GAAGGACAGTGTGAGAGCCACTG AAAGAGCCACAAATTCCTAACAACCTA	FAM-CCCACCAAGAAATCATGTAAGTAATTTTAAAGTCC TET-CCCACCAAGAAATCATGTAAGTAATTTTAAAGTCC	Exploratory-screen

*-hCV assays were obtained from ABI “Assay on Demand” database (<https://www.store.appliedbiosystems.com>). Primer and probe sequences for AoD are not available.

9.1.2. Sarcoidosis patient phenotypes

Sarcoidosis phenotypes fall into two main categories, namely, acute and chronic sarcoidosis. Classification of sarcoidosis patients in this study into a further ten categories involved the following criteria: disease location, symptoms, duration and therapy. 30.77% of patients in the entire sample have an acute form of sarcoidosis with lung involvement (phenotype groups 1 plus 2), whereas chronic form has 61.13% of patients (groups 3 to 6).

Table 16. Sarcoidosis phenotypes

Sarcoidosis phenotype group	Phenotype description	Proportion of patients (%)
1	Acute sarcoidosis (mainly with Loefgren syndrome), without systemic Cortison therapy, to complete or large recovery, possibly relapse after more than 2 years	12.08
2	Acute Sarcoidosis (mainly with Loefgren syndrome), with systemic Cortison therapy, to complete or large recovery, possibly relapse after more than 2 years	18.69
3	Chronic Sarcoidosis (mainly with stealthy beginning), without systemic Cortison therapy, to complete or large recovery	8.44
4	Chronic Sarcoidosis (mainly with stealthy beginning), with systemic Cortison therapy, to complete or large recovery	14.89
5	Chronic Sarcoidosis (mainly with stealthy beginning), with systemic Cortison therapy, with enduring disease activity	33.33
6	Chronic Sarcoidosis (mainly with stealthy beginning), with systemic immunosuppressive therapy, lung transplantation	4.47
7	Sarcoidosis without lung participation (mainly with skin Sarcoidosis)	0.66
8	Sarcoidosis in the infancy	0.16
9	Not classifiable (mainly with missing documents, contradictory information)	3.39
10	Coincidence findings, no complaints	3.89

9.1.3. Structure of 138 sarcoidosis families investigated in the CARD15 study

Table 17. Structure of sarcoidosis families investigated in CARD15 association study

Relationship of patients	No. families	No. patients	Male	Female
Affected siblings				
sister / sister	22	44		44
sister / brother	27	54	27	27
brother / brother	11	22	22	
3 siblings	6	18	9	9
Affected parent / offspring				
mother / daughter	18	36		36
mother / son	8	16	8	8
father / daughter	5	10	5	5
father / son	10	20	20	
Affected 2 and 3 relatives	19	38	21	17
Complex families				
Three patients	7	21	9	12
Four patients	3	12	9	3
Five patients	1	5	1	4
Six patients	1	6	4	2
Total	138	302	135	167

9.1.4. Assays for *CARD15* genotyping

Table 18. SNP assays for *CARD15* genotyping

Marker	Forward primer Reverse primer	Probe 1 Probe 2
SNP5	5'-GGAGGACATATACACAGAGAATGTCC 5'-ATGGCTGGATCCCCGCAGAAGAGCT	FAM-ATGGCTGGACCCCCGCAGAAGAGT TET-ATGGCTGGATCCCCGCAGAAGAGCT
SNP8	5'-TTCCTGGCAGGGCTGTTGTC 5'-AGTGAAGTGCTTGCGGAGG	FAM-CCTGCTCCGGCGCCAGGC TET-CCTGCTCTGGCGCCAGGCC
SNP12	5'-ACTCACTGACACTGTCTGTTGACTCT 5'-AGCCACCTCAAGCTCTGGTG	FAM-TTTTCAGATTCTGGGGCAACAGAGTGGGT TET-TTCAGATTCTGGCGCAACAGAGTGGGT
SNP13	5'-GTCCAATAACTGCATCACCTACCTAG 5'-CTTACCAGACTTCCAGGATGGTGT	FAM-CCCTCCTGCAGGCCCTTGAAA TET-CCTCCTGCAGGCCCTTGAAA

9.1.5. Reaction conditions for diallelic genotyping assays

Table 19. Part 1: reaction conditions of SNP assays in chromosome 6 and CARD15 genotyping experiments

Assay	Forward primer (nmol final concentration)	Reverse primer (nmol final concentration)	Probe 1 (nmol final concentration)	Probe 2 (nmol final concentration)	Annealing temperature (°C)	Number of cycles
rs2073528	900	900	200	200	60	45
rs2072803	900	900	200	200	60	45
hCV2474907	900	900	200	200	60	45
hCV2474883	900	900	200	200	60	45
hCV9804528	900	900	200	200	60	45
rs3736781	900	900	200	200	60	45
rs126007	300	300	100	100	62	45
rs929042	900	900	100	100	62	45
rs53161	300	300	100	100	62	40
rs404240	50	900	100	100	62	45
rs362536	300	900	100	100	60	45
rs1805057	300	300	100	100	60	45
rs3130253	900	900	100	100	62	45
rs1569315	300	300	100	100	64	45
rs8347	300	300	100	100	62	45
rs1264457	300	300	100	100	62	45
rs2074510	300	300	100	100	62	45
rs3095318	300	50	100	100	60	45
rs2073721	50	300	100	100	62	45
rs1050393	300	300	100	100	60	45
rs1051488	900	900	100	100	60	45
rs1063630	900	900	200	200	60	45
rs1063635	300	900	100	100	62	45
rs3828916	900	900	200	200	60	45
rs3131639	900	900	200	200	60	45
rs3134900	300	300	100	100	62	45
AC006046_G30156A	900	900	200	200	60	45
rs3130062	900	300	100	100	62	40
rs1800683	900	900	200	200	60	45
rs1041981	900	900	100	100	65	40

Table 19. Part 2: reaction conditions of SNP assays in chromosome 6 and CARD15 genotyping experiments

Assay	Forward primer (nmol final concentration)	Reverse primer (nmol final concentration)	Probe 1 (nmol final concentration)	Probe 2 (nmol final concentration)	Annealing temperature (°C)	Number of cycles
rs1799964	300	300	100	100	62	55
rs1799742	300	300	200	200	60	50
rs1800629	900	300	100	100	62	40
rs361525	900	300	100	100	62	40
rs707916	300	300	100	100	62	45
rs707926	50	900	100	100	60	45
rs2075800	300	300	100	100	60	45
rs539689	50	900	100	100	60	45
AF019413_C89532G	300	900	100	100	62	45
rs1035798	50	900	100	100	60	45
rs2070600	900	900	200	200	60	45
rs1800624	900	900	200	200	60	45
rs2071287	900	900	200	200	60	45
rs384247	900	900	200	200	60	45
rs915894	900	900	200	200	60	45
rs436845	900	900	200	200	60	45
rs560505	900	900	200	200	60	45
rs485774	900	900	200	200	60	45
rs552339	900	900	200	200	60	45
rs761187	900	900	200	200	60	45
rs2073044	900	900	200	200	60	45
rs2050190	900	900	200	200	60	45
rs2073045	900	900	200	200	60	45
AL034394_T26355G	900	900	200	200	60	45
AL034394_A26270C	900	900	200	200	60	45
AL034394_C20283T	900	900	200	200	60	45
AL034394_C16965T	900	900	200	200	60	45
hCV2488476	900	900	200	200	60	45
AL034394_C12339T	900	900	200	200	60	45
AL034394_T10552C	900	900	200	200	60	45
BTNL2_E360Amb	900	900	200	200	60	45

Table 19. Part 3: reaction conditions of SNP assays in chromosome 6 and CARD15 genotyping experiments

Assay	Forward primer (nmol final concentration)	Reverse primer (nmol final concentration)	Probe 1 (nmol final concentration)	Probe 2 (nmol final concentration)	Annealing temperature (°C)	Number of cycles
BTNL2_P299Q	900	900	200	200	60	45
BTNL2_M286I*	900	900	200	200	60	45
BTNL2_P285L	900	900	200	200	60	45
rs2076530	900	900	200	200	60	45
BTNL2_S240L	900	900	200	200	60	45
rs2076529	900	900	200	200	60	45
rs2294878	900	900	200	200	60	45
BTNL2_i4_A239T	900	900	200	200	60	45
rs2076524	900	900	200	200	60	45
BTNL2_i3_C1900T	900	900	200	200	60	45
rs2076523	300	300	100	100	62	45
BTNL2_i3_A1837G	900	900	200	200	60	45
BTLIIW94R	900	900	200	200	60	45
BTLIIH60H	900	900	200	200	60	45
BTNL2_i1_G1717T	900	900	200	200	60	45
hCV2455668	900	900	200	200	60	45
DRAp_231	900	900	200	200	60	45
DRAp_224	900	900	200	200	60	45
DRAp_196	900	900	200	200	60	45
DRAp_127	900	900	200	200	60	45
rs14004	900	900	200	200	60	45
hCV2455646	900	900	200	200	60	45
rs8084	900	900	200	200	60	45
rs7192	300	300	100	100	62	45
rs2213584	900	900	200	200	60	45
rs2395182	900	900	200	200	60	45
rs707952	300	300	100	100	62	45
X66401_G48629A	300	300	100	100	62	45
rs1042337	900	900	100	100	62	45
rs1063478	300	300	100	100	62	45
rs3918143	300	300	100	100	64	45

Table 19. Part 4: reaction conditions of SNP assays in chromosome 6 and CARD15 genotyping experiments

Assay	Forward primer (nmol final concentration)	Reverse primer (nmol final concentration)	Probe 1 (nmol final concentration)	Probe 2 (nmol final concentration)	Annealing temperature (°C)	Number of cycles
rs2581	50	300	100	100	62	45
rs926424	900	50	100	100	60	45
rs1547387	50	900	100	100	62	45
rs466384	900	50	100	100	60	45
rs2071888	300	900	100	100	62	45
rs999942	50	900	100	100	60	45
rs943461	50	900	100	100	62	45
rs755658	300	300	100	100	62	45
rs992105	300	300	100	100	60	45
rs1320991	300	300	100	100	62	45
rs1326752	900	300	100	100	62	45
Z95152_G57723A	50	300	100	100	62	45
Z95152_T58784G	300	300	100	100	62	45
Z95152_T106861C	50	900	100	100	62	45
rs763096	50	300	100	100	62	45
rs2007683	300	900	100	100	62	45
rs881648	300	900	100	100	62	45
rs881647	900	300	100	100	60	45
rs916287	50	300	100	100	62	45
rs1885205	50	300	100	100	62	45
rs2055272	50	300	100	100	62	45
rs1004279	50	900	100	100	60	45
rs1541816	300	900	100	100	62	45
rs1014737	300	300	100	100	62	45
rs2234246	300	300	100	100	64	40
rs1817538	50	300	100	100	60	45
rs1351835	900	300	100	100	62	45
rs2234237	50	300	100	100	62	45
AL391903_C88694T	900	900	100	100	62	45
rs742735	300	900	100	100	60	45
rs1410492	50	900	100	100	60	45

Table 19. Part 5: reaction conditions of SNP assays in chromosome 6 and CARD15 genotyping experiments

Assay	Forward primer (nmol final concentration)	Reverse primer (nmol final concentration)	Probe 1 (nmol final concentration)	Probe 2 (nmol final concentration)	Annealing temperature (°C)	Number of cycles
rs1160867	50	900	100	100	58	45
rs9395	300	300	100	100	60	45
rs1418488	50	300	100	100	62	45
NOD2_SNP5	900	900	100	100	60	45
NOD2_SNP8	400	400	100	100	62	45
NOD2_SNP12	300	300	100	100	62	45
NOD2_SNP13	400	400	100	100	62	45

*-all four probes of the assay had the same 200nmol concentration

9.1.6. Mutation detection primers used for chromosome 6 PCR and sequencing experiments

Table 20. Part1: mutation detection primers and SNP markers selected for chromosome 6 high density LD mapping experiment.

Primer	Sequence	Ampl icon length	SNP marker selected for genotyping*
1_For 1_Rev	5'-CCTCACACATGCCAATTCCC-3' 5'-TTTTGCCTCCCCATCGAC-3'	463bp	AL034394_T26355G AL034394_A26270C
2_For 2_Rev	5'-CCTCACACATGCCAATTCCC-3' 5'-TTTTGCCTCCCCATCGAC-3'	359bp .	AL034394_C20283T
3_For 3_Rev	5'-CAAACCACCACTCTCTTTCATTG-3' 5'-CTCCCTCTTGTTTTTCTCCATTCT-3'	384bp .	AL034394_C16965T
4_For 4_Rev	5'-TAAACATTTGTGGAAGAATAATGTGTAGTT-3' 5'-TCAAGAAGATAAACTATATGTGGAAGACAG-3'	271bp .	AL034394_C12339T
5_For 5_Rev	5'-GAGCTGTGTTTGAAGCCAATGTC-3' 5'-GAGCCTCCCTTTTCACAAGTCA-3'	540bp .	AL034394_T10552C
BTNL2_e8_For BTNL2_e8_Rev	5'-CTTGTCTTCCTGTTTGGCTGTAA-3' 5'-GGGAAGGTGTAGATAGGGCACTT-3'	497bp .	
BTNL2_e7_For BTNL2_e7_Rev	5'-AGCAGGCCAACTGCTTCCT-3' 5'-ACTGAGCCTGGATTGCATGAT-3'	402bp .	
BTNL2_e6_For BTNL2_e6_Rev	5'-CCTGGCCGAAGTTATTTTGT-3' 5'-TAGAATCCCTGGGTGTCCTG-3'	738bp .	BTNL2_e6_E360Amb BTNL2_e6_P299Q BTNL2_e6_M286I BTNL2_e6_P285L
BTNL2_e5_For BTNL2_e5_Rev	5'-GGTTTCTAAACTCCAATGGAGCTGTT-3' 5'-CAAATGTCAGAGAAATTGTCCAGGA-3'	528bp .	BTNL2_e5_S240L
BTNL2_i4_For BTNL2_i4_Rev	5'-GGTAAGTACGAGGTGCTGGCA-3' 5'-TTGCTACACTCACAGTATCCCAGG-3'	541bp .	BTNL2_i4_A239T
BTNL2_e4_For BTNL2_e4_Rev	5'-GTGACAGCATTTTTTGTGTCAG-3' 5'-GGTGCCATCTGATGCTCACTG-3'	474bp .	
BTNL2_i3_For BTNL2_i3_Rev	5'-CCTCAGACTCCTGAACTCCG-3' 5'-GGTGCTATGGTGCAGTCCC-3'	469bp .	BTNL2_i3_C1900T BTNL2_i3_A1837G
BTNL2_e3_For BTNL2_e3_Rev	5'-CATTTTGACAGGCTGGACACC-3' 5'-CCTCAACTGTCACAAAGCTTACCA-3'	511bp .	BTNL2_e3_W94R BTNL2_e3_H60H

Table 20. Part2: mutation detection primers and SNP markers selected for chromosome 6 high density LD mapping experiment.

Primer	Sequence	Ampl icon length	SNP marker selected for genotyping*
BTNL2_e2_For BTNL2_e2_Rev	5'-CAATGGAAGCCATGGAGTGTG-3' 5'-AACGTGGTGCAACCACTGATT-3'	489bp .	
BTNL2_i1_For BTNL2_i1_Rev	5'-AAGCCGAAC TACATCAACGTTG-3' 5'-CTGCACCAGACAGATTGTAGCC-3'	528bp .	BTNL2_i1_G1717T
BTNL2_e1_For BTNL2_e1_Rev	5'-TGTCTCTAAGGGCAAACAGC-3' 5'-GTTTCCCCCTTTCCAAGAACTA-3'	693bp .	
DRAp_For DRAp_Rev	5'-TAGGTGTTCCATTGATTCTATTCTCAC-3' 5'-TGATAGCCCATGATTCCTGAGC-3'	462bp .	06DRAp_231 06DRAp_224 06DRAp_196 06DRAp_127
DRA_5'_For DRA_5'_Rev	5'-CCCAAGCCTAACTCATTTCAACA-3' 5'-CCCCAAGGTATGGATCTCAGAAT-3'	491bp .	rs2213584 rs2395182

*- only markers selected for genotyping are represented in the table. Data of other SNPs is not shown.

9.1.7. Sequence variations detected via sequencing

New sequence variations observed in the mutation detection experiment and selected for the further high LD mapping experiment:

AL034394_T26355G

5'_FLANK:ATCAGGGAAAGAAAATGTCTATGGCAAAGGGCCTAAGATTAGAATC

OBSERVED:T/G

3'_FLANK:ACTACGCTCAGCGAAGTTGCACTCTCATCTAAGAAGTTCTTATCTAAA

AL034394_A26270C

5'_FLANK: AGCCACGTAGGCTTGAGACACAGATTCTGGAACCTGAGGGAAAGAAA

OBSERVED:A/C

3'_FLANK:ATGTGTTTGCAGATATAAAAATATGGATTTAAAAAAATCATCAGGGAAA

AL034394_C20283T

5'_FLANK:ATCTGAACTAAGAGATGACAAGAAAGAGGGGTACGGCAGAAATAAG

OBSERVED:C/T

3'_FLANK:AATAGCCCTGCCAGTACACAGTGTCTTAAGCTCCTACTTCTTGATTTTT

AL034394_C16965T

5'_FLANK:CCTTGTCTTGGTGTGTGGGGGCATCCCAGTTCGGGGAACATTCTC

OBSERVED:C/T

3'_FLANK:GTTGAAGAAGTGTATCTTCTACCCTCAGATCGGATGAGGTTGGGATGA

AL034394_C12339T

5'_FLANK:ACGCTCTGCTGCTCTCCAGGATGGCAGACGTCAGAGGCATCGCGGTCC

OBSERVED:C/T

3'_FLANK:ATGCGGATCACTCATGAACAGCACTCTGAACTTGCTCTTCTGGGCCAT

AL034394_T10552C

5'_FLANK:GTTAGGGAGGCTGTTTTGTGTGCTCTAGAAAGTGCTTCCACTTTAAATTTCTTA

OBSERVED:T/C

3'_FLANK:GTCCGTGAAACACTTGTCTTCTGTTTGGCTGTTAATAGAATGAAATTATATTG

BTNL2_e6_E360Amb

5'_FLANK:ATCAGCATCCCCTTTTTGGGCGAGGAGAAAATCGCAACTTTTTCTCTCTCA

OBSERVED:G/T

3'_FLANK:GTTGGTGATTTCCTTATGTTCCCTTCAGGTTTGGAAAATAAATATGAAGACCAACT

BTNL2_e6_P299Q

5'_FLANK:AGAAATGCAGCCGATGTGCTCTTCAGATGGGTGGTTCCCACAGCCCCACGTGC

OBSERVED:CA/AG

3'_FLANK:TGGAGGGACATGGAAGGAAAGACGATACCATCATCTTCCCAGGCCCTGACTC

BTNL2_e6_M286I

5'_FLANK:CTTCCCCACTGATCACTGTGGAGGGGCAAGAAGATGGAGAAATGCAGCCGAT

OBSERVED:G/A

3'_FLANK:TGCTCTTCAGATGGGTGGTTCCCACAGCCCCACGTGCCATGGAGGGACATGGA

BTNL2_e6_P285L

5'_FLANK:GGTTCTTCCCCACTGATCACTGTGGAGGGGCAAGAAGATGGAGAAATGCAGC

OBSERVED:C/T

3'_FLANK:GATGTGCTCTTCAGATGGGTGGTTCCCACAGCCCCACGTGCCATGGAGGGACA

BTNL2_e5_S240L

5'_FLANK:TGACGCCATTGACGAGGGCAGACTGACCCTGCAGATACTCAGTGCCAGACCTT

OBSERVED:C/T

3'_FLANK:GGACGACGGGCAGTACCGCTGCCTTTTTTGAAAAAGATGATGTCTACCAGGAGG

BTNL2_i4_A239T

5'_FLANK:TGTGTTACCTGACACTTTGGGGGCTCATGGAGATTCCAGTGAGCATCAGATGG

OBSERVED:A/T

3'_FLANK:CACCTATAGGAAGCATGGCATCATCTCTGTTCTATGTGTTGAGGGGAAAGGGA

BTNL2_i3_C1900T

5'_FLANK:GCTGCTGGCCGTGTCTGAGCATCGCATCCAAGATAAAGATGGCCTGTTCTATG

OBSERVED:C/T

3'_FLANK:GGAAGCCACCCTGGTGGTCAGGAACGCCTCTGCAGAGTCTGTGTCCTGCTTGG

BTNL2_i3_A1837G

5'_FLANK:GTGCACTGCAAGGGGCTGGTTCCCAGAGCCCCAGGTGTATTGGGAAGACATCC

OBSERVED:G/A

3'_FLANK:GGGAGAGAAGCTGCTGGCCGTGTCTGAGCATCGCATCCAAGATAAAGATG

BTNL2_e3_W94R

5'_FLANK:GCACAGGGATGGAGTGGAGGTGACTGAGATGCAGATGGAGGAGTACAGAGGC

OBSERVED:T/A

3'_FLANK:GGGTAGAGTGGATAGAGAATGGCATTGCAAAGGGAAATGTGGCACTGAAGAT

BTNL2_e3_H60H

5'_FLANK:GGGAAGATGCCCTGTAACTGCCAGCTACTCCCAAGAGGACCACAATGCA

OBSERVED:C/T

3'_FLANK:GTGGAGGTGAGGTGGTACCGCTCAGAGCCCAGCACACCTGTGTTTGTGCACAG

BTNL2_i1_G1717T

5'_FLANK:GGATGTGATCAGTAATTGCAGAGAAATGGAGCCAAGCCCATTTCACTGACTGT

OBSERVED:G/T

3'_FLANK:CAGGGGTTGATGTGGATTCTCTGCTAAGACTGCATAACCAATGGAAGCCATGG

HLA-DRA-G231C

5'_FLANK:ACAAACAGGACAACAACGACAACAAATCAAGTATTTAAGACTGTTGGTCAAT

OBSERVED:G/C

3'_FLANK:ACGGACAGACTCTTTAACCCCAATCTGAGTTCTAGACAGGGATATACCTGAAG

HLA-DRA-G224A

5'_FLANK:GTTCTTAAACAAACAGGACAACAACGACAACAAATCAAGTATTTAAGACTGTT

OBSERVED:G/A

3'_FLANK:GTCAATGACGGACAGACTCTTTAACCCCAATCTGAGTTCTAGACAGGGATATA

HLA-DRA-G196A

5'_FLANK:TACTCCGTTTCATTGGATAAAGAAGTAAAGTTCTTAAACAAACAGGACAACAAC

OBSERVED:G/A

3'_FLANK:ACAACAAATCAAGTATTTAAGACTGTTGGTCAATGACGGACAGACTCTTTAAC

HLA-DRA-G127A

5'_FLANK:GAAAAATATTTTGAGATGACGCATCTGTTGCTAGGGGAAGGGTTCTTGCAAAG

OBSERVED:G/A

3'_FLANK:GTCCAGGACACAAGATACTCCGTTTCATTGGATAAAGAAGTAAAGTTCTTAAAC

Frequencies of listed variations in control population:

AL034394_T26355G:T=0.93/G=0.07
AL034394_A26270C:A=0.74/C=0.26
AL034394_C20283T:C=0.97/T=0.03
AL034394_C16965T:C=0.85/T=0.15
AL034394_C12339T:C=0.86/T=0.14
AL034394_T10552C:T=0.86/C=0.14
BTNL2_e6_E360Amb:G=0.996/T=0.004
BTNL2_e6_P299Q:CA=0.85/AG=0.15
BTNL2_e6_M286I:G=0.85/A=0.15
BTNL2_e6_P285L:C=0.85/T=0.15
BTNL2_e5_S240L:C=0.98/T=0.02
BTNL2_i4_A239T:A=0.97/T=0.03
BTNL2_i3_C1900T:C=0.92/T=0.08
BTNL2_i3_A1837G:G=0.92/A=0.08
BTNL2_e3_W94R:T=0.92/A=0.08
BTNL2_e3_H60H:C=0.92/T=0.08
BTNL2_i1_G1717T:G=0.98/T=0.02
HLA-DRA-G231C:G=0.30/C=0.70
HLA-DRA-G224A:G=0.30/A=0.70
HLA-DRA-G196A:G=0.31/A=0.69
HLA-DRA-G127A:G=0.98/A=0.02

9.1.8. Sequence data

Stammers et al. ¹⁴⁴ initially described BTL-II sequence information. Six separate entries in the GenBank corresponded to each exon: AF186588, AF186589, AF186590, AF186591, AF186592 and AF186593, all located in the BAC insert AL935032.

The revised structure of the BTNL2 gene comprises 8 exons:

Exon 1

ACCCTTCAGTCTAAGGGGAGGTGACTGGTAAGTTTCAG

Exon 2

GGAGGATGGTGGATTTCCAGGCTACAATCTGTCTGGTGCAGTCGCCTCCTCCTATTTCATCCTGCT
GACAATGAAGCAGTCAG

Exon 3

AAGACTTTAGAGTCATTGGCCCTGCTCATCCTATCCTGGCCGGGGTTGGGGAAGATGCCCTGTAA
CCTGCCAGCTACTCCCCAAGAGGACCACAATGCACGTGGAGGTGAGGTGGTACCGCTCAGAGCCC
AGCACACCTGTGTTTGTGCACAGGGATGGAGTGGAGGTGACTGAGATGCAGATGGAGGAGTACAG
AGGCTGGGTAGAGTGGATAGAGAATGGCATTGCAAAGGAAATGTGGCACTGAAGATACACAACA
TCCAGCCCTCCGACAATGGACAATACTGGTGCCATTTCCAGGATGGGAACTACTGTGGAGAAACAA
GCTTGCTGCTCAAAGTAGCAG

Exon 4

AGAAACTCCAGACTGAGCTGG

Exon 5

CTTCTTTAAAAGTGAATGGACCTTCCCAGCCCATCCTCGTCAGAGTGGGAGAAGATATACAGCTAA
CCTGTTACCTGTCCCCAAGGCGAATGCACAGAGCATGGAGGTGAGGTGGGACCGATCCCACCGTT
ACCCTGCTGTGCATGTGTATATGGATGGGGACCATGTGGCTGGAGAGCAGATGGCAGAGTACAGA
GGGAGGACTGTACTGGTGTAGTGACGCCATTGACGAGGGCAGACTGACCCTGCAGATACTCAGTGC
CAGACCTTCGGACGACGGGCAGTACCGCTGCCTTTTTGAAAAGATGATGTCTACCAGGAGGCCAG
TTTGATCTGAAGGTGGTAG

Exon 6

GTCTGGGTTCTTCCCCTGATCACTGTGGAGGGGCAAGAAGATGGAGAAATGCAGCCGATGTGCT
CTTCAGATGGGTGGTTCCCACAGCCCCACGTGCCATGGAGGGACATGGAAGGAAAGACGATACCA

TCATCTTCCCAGGCCCTGACTCAAGGCAGCCACGGGCTGTTCCACGTGCAGACATTGCTAAGG
 GTCACAAACATCTCCGCTGTGGACGTCACCTGTTCCATCAGCATCCCCTTTTTGGGCGAGGAGAAA
 ATCGCAACTTTTTCTCTCTCAG

Exon 7

AGTCCAGGATGACATTTTTGTGGAAAACACTGCTTGTGGGGATTGCTTCTTGCTGTGGCTGTAGG
 CCTGCCAGGAAGAGGAGCTGAAAAGAG

Exon 8

TGAATGTGACATTGGCTTCAAACACAGCTCACCTGAGACTGATTTCTTCTGAACAGAACAAGCGTG
 TGATCCATGGACATTCAGGCAGCCAGATATCCCACAGAGATCTGACTATCTGCTCTATGGGC

The protein structure of the revised BTNL2 has two splice variants.

The long splice form of the BTNL2 protein:

MVDFPGYNLSGAVASFLFILLTMKQSEDFRIGPAHPILAGVGEDALLTCQLLPKRRTTMHVEVRWYRSE
 PSTPVFVHRDGVEMTEMQMEEYRGWVEWIENGIKGNVALKIHNIQPSDNGQYWCHFQDGNYCGETS
 LLLKVAEKLQTELASLKVNQPSQPILVRVGEDIQLTCYLSPKANAQSMEVRWDRSHRYPAVHVYMDG
 DHVAGEQMAEYRGRTVLVSDAIDEGRLTLQILSARPSDDGQYRCLFEKDDVYQEASLDLKVGLGSSP
 LITVEGQEDGEMQPMCSSDGWFPQHPVWRDMEGKTIPSSQALTQGSGLFHVQTLRVTNISAVDV
 TCSISIPFLGEEKIATFSLSESMTFLWKTLVWGLLLAVAVGLPRKRS

The short splice form of the BTNL2 protein:

MVDFPGYNLSGAVASFLFILLTMKQSEDFRIGPAHPILAGVGEDALLTCQLLPKRRTTMHVEVRWYRSE
 PSTPVFVHRDGVEMTEMQMEEYRGWVEWIENGIKGNVALKIHNIQPSDNGQYWCHFQDGNYCGETS
 LLLKVAEKLQTELASLKVNQPSQPILVRVGEDIQLTCYLSPKANAQSMEVRWDRSHRYPAVHVYMDG
 DHVAGEQMAEYRGRTVLVSDAIDEGRLTLQILSARPSDDGQYRCLFEKDDVYQEASLDLKVWVWLP

9.2. Additional experiments on BTNL2

This chapter includes unpublished (to this date) laboratory experiments performed with collaborators: Dr.K.A.Gaede (BAL extraction), Dr.P.Rosenstiel Dr.K.Huse cDNA (cloning and expression experiments), Mario Albrecht (*in silico* analysis).

9.2.1. Preparation of Bronchoalveolar lavage cell samples

Bronchoalveolar lavage (BAL) cell samples were taken from eight chronic sarcoidosis patients with overt disease (average age: 44.8 ± 3.6 years), and four controls (average age: 32.7 ± 8.2 years). All sarcoidosis patients showed clinical signs of active disease. Four patients, who underwent bronchoscopy for evaluation of chronic coughing and were retrospectively found to be free of inflammatory or malignant disorder, served as additional controls. All patients and controls gave written informed consent to the study. The protocol was approved by the institutional review board of the institution. BAL was performed as previously described¹⁷⁵. Mean recovery was 62.5% and the average cell number/100ml of BAL cells was $12.9 \pm 3.3 \times 10^6$. The cells were centrifuged at 500 x g and washed three times with phosphate buffered saline (PBS) at +4°C. The cell suspensions used in this study contained $85.4 \pm 5.3\%$ alveolar macrophages. Total cellular RNA was extracted from BAL cells using TRIzol reagent® (Invitrogen, Paisley, UK) according to the manufacturer's recommendations. To remove contaminating genomic DNA, total RNA samples were treated with deoxyribonuclease I (Boehringer Mannheim, Mannheim, FRG). Reverse transcription was performed on equal amounts of total RNA per sample, using 1x First Strand Buffer (Invitrogen, Paisley, UK), 2'-deoxynucleotide 5'triphosphate (Invitrogen, Paisley, UK), dithiothreitol, oligo (dT)12-18 primer (Invitrogen, Paisley, UK), RNasin, and Superscript RT reverse transcriptase (RT) (Invitrogen, Paisley, UK). Prior to addition of the RT, the RT mixture was incubated for 5 min at 65°C, followed by an incubation period of 5 min at 37°C. After addition of the enzyme, incubation was continued for another 1 h at 37°C. For quality control, each RNA sample was subjected to RT-PCR omitting the RT.

9.2.2. *BTNL2* expression experiment

In order to verify the published *BTNL2* gene model, conserved exons were searched in mouse and human genomic sequences (Twinscan gene predictions, genome.ucsc.edu) and related to predicted splice sites (www.fruitfly.org/seq_tools/splice.html). The predicted exons were verified by rt-PCR in a pool of cDNA tissue samples (Human Multiple Tissue cDNA Panels I and II, Clontech; data not shown) and by subsequent cloning and sequencing of PCR products. Primers for rt-PCR and cDNA cloning are listed in table 21.

Table 21. Primers for rt-PCR and cDNA cloning. The position of the primer sequences in the revised *BNTL2* transcript refers to the 5' end for forward primers and to the 3' end for reverse primers.

Primer	Sequence	Position in cDNA
BTNL2f1	ACCCTTCAGTCTAAGGGGAG	1
BTNL2f2	TGGTGGATTTTCCAGGCTAC	45
BTNL2f3	TTAGAGTCATTGGCCCTGCT	126
BTNL2r1	CTTCAGTGCCACATTTCCCT	370
BTNL2r2	CTTCTCCCACTCTGACGAGG	542
BTNL2f4	CCTGCAGATACTCAGTGCCA	733
BTNL2r3	CGAAGGTCTGGCACTGAGTA	760
BTNL2f5	AGTACCGCTGCCTTTTTGAA	771
BTNL2r4	TTCAAAAAGGCAGCGGTACT	790
BTNL2f6	TCGTCTTTCCTTCCATGTCC	958
BTNL2f7	CGGAGATGTTTGTGACCCTT	1042
BTNL2f8	ACTTGTTCCATCAGCATCCC	1073
BTNL2r5	TGGATCACACGCTTGTTCTG	1282
BTNL2r6	GCCCATAGAGCAGATAGTCAG	1337

For expression analysis, nested PCR in THP-1 cells and BALs was performed with primers 5'-TTAGAGTCATTGGCCCTGCT and 5'-GCCCATAGAGCAGATAGTCAG for the first, and 5'-AGTACCGCTGCCTTTTTGAA and 5'-TGGATCACACGCTTGTTCTG for the second round of amplification.

9.2.3. Verification of the splicing effects of rs2076530

Matching pairs of cDNA and DNA from lymphoblastoid cell lines (upper part) and from the peripheral blood of normal controls (lower part) were genotyped and the *BTNL2* splice forms evaluated by rtPCR, cloning and sequencing of PCR products.

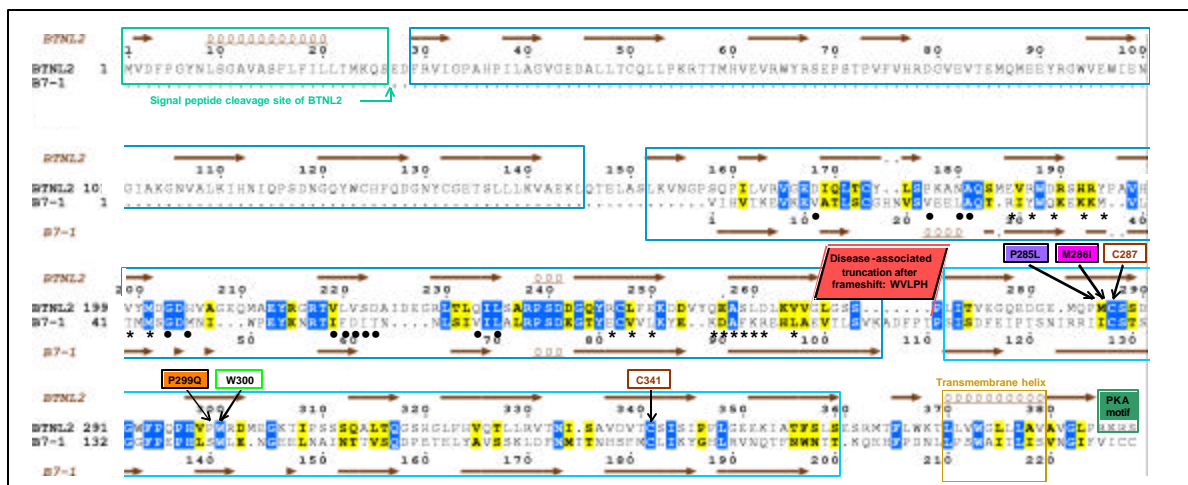
Table 22. *BTNL2* rs2076530 genotype correlation with two different splicing forms.

Source	rs2076530 genotype	no. samples	no. clones	4 bp present	4 bp deleted
cell lines	GG	4	64	64	0
	AG	5	98	22	76
	AA	3	131	0	131
leucocytes	GG	1	36	36	0
	AG	3	36	9	27
	AA	3	35	0	35

9.2.4. In silico *BTNL2* protein analysis

Domain architecture of the *BTNL2* gene product was derived computationally. The figure 17 shows a sequence-structure alignment of *BTNL2* to the IgV domain and the following IgC domain of B7-1 (PDB identifier 1i8l). Coloured frames embrace the putative N-terminal signal peptide (light green), the three immunoglobulin-like domains IgV-IgV-IgC (blue), and a potential C-terminal PKA phosphorylation motif (dark green). The DSSP secondary structure of 1i8l and the corresponding predictions by the PSIPRED server for *BTNL2* are depicted on top and below each alignment row, respectively (curled lines: α -helices, horizontal arrows: β -strands). The predicted transmembrane helices are annotated at the C-termini. Blue and yellow boxes highlight identical and conserved residues, respectively. The disease-associated sequence truncation of *BTNL2*, variants P285L, M286I and P299Q, a strictly conserved tryptophan W300, and the adjacent disulfide bond between residues C287 and C341 are annotated by text labels. Asterisks in the IgV domain of B7-1 indicate residues involved in binding to the CTLA-4 receptor; solid circles mark residues participating in the dimer interface of B7-1.

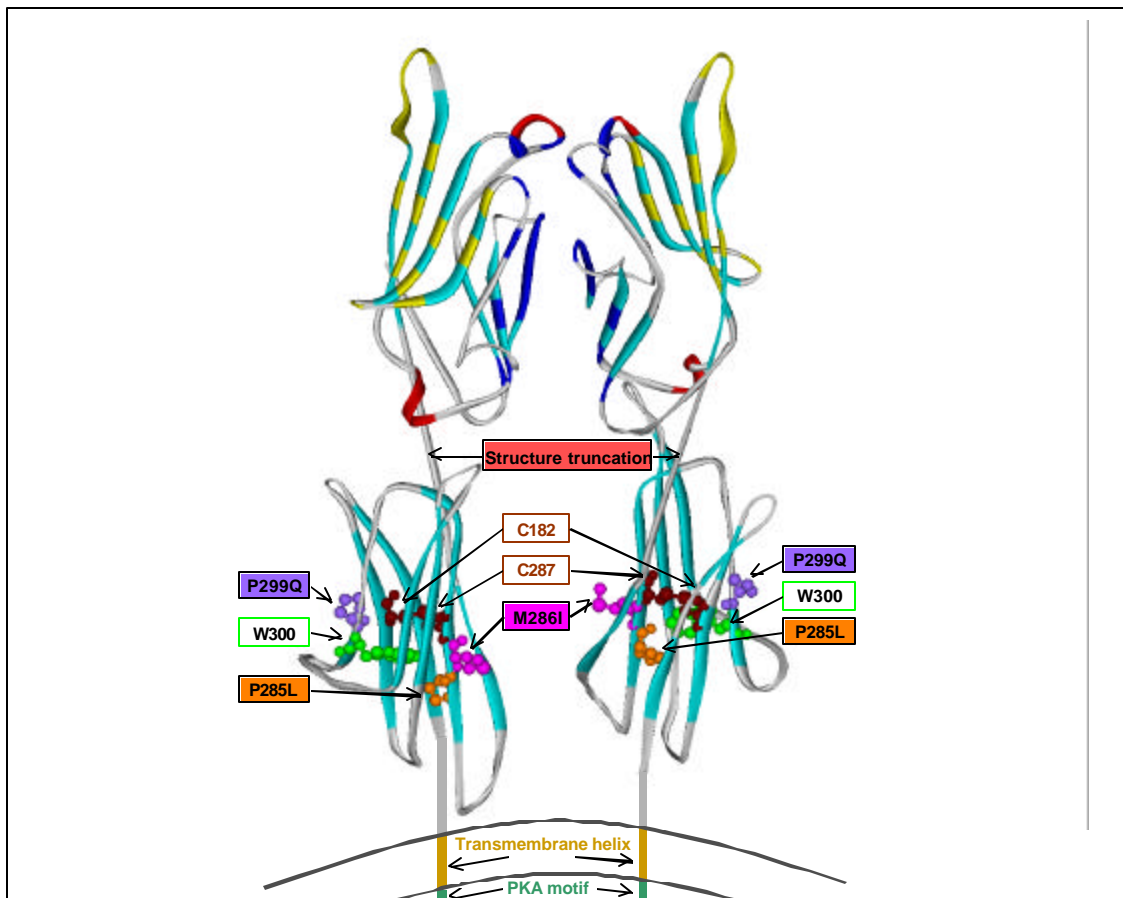
Fig.17. In silico *BTNL2* protein alignment to the B7-1 molecule



9.2.5. Computational 3D structure model of the second IgV domain and the following IgC domain of the BTNL2 homodimer

3D structure model of the second IgV domain (upper half) and the following IgC domain (lower half) of the BTNL2 homodimer were derived computationally. The model was created using the B7-1 structure as a template. The location of the C-terminal truncation and of sequence variants P285L, M286I and P299Q are annotated. A stabilising disulfide bond between residues C287 and C341, which are located adjacent to variants P285L and M286I, is highlighted in brown. A strongly conserved tryptophan W300 located close to variant P285L and the disulfide bond are highlighted in light green. The putative binding site of BTNL2 to a yet unknown receptor is marked in the IgV domain in yellow. This site has been deduced from the location of residues interacting with the B7-1 receptor CTLA-4. Residues participating in the dimerisation interface of the IgV domain are shown in dark blue. The figure also indicates the predicted transmembrane helix and a potential PKA phosphorylation site at the C-terminus of BTNL2.

Fig.18. Second IgV domain and the following IgC domain of the BTNL2 homodimer



Index of figures

	Page
Fig.1. Chromosome 6 multipoint NPL results	17
Fig.2. The Major histocompatibility complex (published by ⁶³)	18
Fig.3. Gene content in peri-MHC region (published by ⁸⁰)	22
Fig. 4. Schematic view of TaqMan® assay	36
Fig.5. TaqMan® assay using four-probe combination	38
Fig.6. Schematic representation of fluorescent measurement of TaqMan® assay	39
Fig.7. Fluorescence reading of a 96-well plate for the amplification group 1	45
Fig.8. Screenshot of the "DRB tool" software	47
Fig.9. Graphical representation of the stage I SNP screen on chromosome 6p21	58
Fig10. Linkage disequilibrium between 69SNP markers typed for association screening on chromosome 6	63
Fig.11. Fragment of the high density SNP map establishment in the association locus	65
Fig.12.High density mapping in the association region	66
Fig.13. The matrix of the r^2 measured in the locus <i>AGER-HLA-DQA1</i> (from rs1035798 to rs707952 used in the stage II chromosome 6 screening)	67
Fig.14. Gene content in the fragment of the chromosome 6p21	70
Fig.15. Expression pattern of the <i>BTNL2</i> gene	72
Fig.16. Alternative splicing of the <i>BTNL2</i> transcript at exon 5	74
Fig.17. In silico <i>BTNL2</i> protein alignment to the B7-1 molecule	132
Fig.18. Second IgV domain and the following IgC domain of the <i>BTNL2</i> homodimer	133

Index of tables

	Page
Table 1. Materials	26
Table 2. Equipment	28
Table 3. Sarcoidosis sample for chromosome 6 association mapping study	31
Table 4. Known <i>HLA-DRB1</i> genotypes of patients used as positive controls for amplification of <i>HLA-DRB1</i> allele-groups 1 to 7.	35
Table 5. Primers used for the amplification of <i>HLA-DRB1</i> allele groups	42
Table 6. PCR conditions used for <i>DRB1</i> allele-group amplification	43
Table 7. PCR temperature profiles used for <i>DRB1</i> allele-group amplification.	44
Table 8. Analysis of 69SNP marker covering chromosome 6 linkage region typed in the „exploratory” sample	59
Table 9. Significance of disease-associated <i>BTNL2</i> markers in replication sample	68
Table 10. Association of the rs2076530 with susceptibility to sarcoidosis	69
Table 11. Association between <i>BTNL2</i> and <i>DRB1</i> risk alleles and sarcoidosis (based upon 947 cases and 517 controls)	71
Table 12. Differentiation between <i>BTNL2</i> and <i>DRB1</i> effects	71
Table 13. <i>CARD15</i> allele (A) and genotype(B) frequencies in sarcoidosis patients and healthy controls	76
Table 14. Results of transmission disequilibrium test	77
Table 15. Primers and probes of SNP assays	109
Table 16. Sarcoidosis phenotypes	113
Table 17. Structure of sarcoidosis families investigated in <i>CARD15</i> association study	114
Table 18. SNP assays for <i>CARD15</i> genotyping	115
Table 19. Reaction conditions of SNP assays in chromosome 6 and <i>CARD15</i> genotyping experiments	116
Table 20. Mutation detection primers and SNP markers selected for chromosome 6 high density LD mapping experiment	121
Table 21. Primers for rt-PCR and cDNA cloning	130
Table 22. <i>BTNL2</i> rs2076530 genotype correlation with two different splicing forms	131

10. Curriculum vitae

PERSONAL INFORMATION

Name: Ruta Valentonyte
 Date of birth: 22nd October 1976
 Place of birth: Panevezys, Lithuania
 Citizenship: Lithuanian
 Family status: Single

EDUCATION

June 1994 General qualification for university entrance; Panevezys secondary school 7, Panevezys, Lithuania

June 1998 Bachelor of natural sciences (biology), Vilnius Pedagogical University, Vilnius, Lithuania.

June 2000 Master of Natural Sciences (genetics). Studies of the following subjects: Medicine Genetics; Mutagenesis; Molecular Cytogenetics; Methods of Molecular Genetics; Genotoxycology; Cytogenetics; Genetics of Somatic Cells; Oncogenetics; Recombinogenesis; Evolutionary Genetics; Molecular Genetics; Genetics of Microorganisms, Vilnius University, Vilnius, Lithuania.

since February 2001 PhD (biology) student at Christian-Albrechts-University Kiel, Germany. Courses during the study period: „Epidemiology and Population Genetics” and „Biometry” courses at the Institute of Medicine Informatics and Statistics, Kiel, Germany; „Advanced Gene Mapping\Linkage Course” organised by Rockefeller University Laboratory of Statistical Genetics, Munich, Germany.

RESEARCH EXPERIENCE

- January 1997-June 1998 Laboratory of Cell Engineering, Institute of Botanic, Vilnius, Lithuania.
Bachelor thesis: „Identification of somatic hybrids of potatoe *Solanum tuberosum* L. and *S.polyadenium* Greenman”.
- July 1998 -2000 June Laboratory of Cell Engineering, Institute of Botanic, Vilnius, Lithuania.
Master thesis: „Expression of the *taqIR* gene in Transgenic Plants”.
- July 2000-January 2001 Research assistant at „Laboratory of Cell Engineering” at the „Institute of Botanic” Vilnius, Lithuania.
- since February 2001 Mucosal immunology group, Department of General Internal Medicine, University Clinic Schleswig-Holstein, Campus Kiel, Germany. PhD student in biology.

11. Declaration (Erklärung) and publication list

I acted as a primary scientist in the positional cloning experiment on chromosome 6. The project has been performed in collaboration between Lübeck (Dr.M.Schürmann¹), Borstel (Dr.K.A.Gaede²), Kiel (Dr.P.Rosenstief³) and Jena (Dr.K.Huse⁴). The collaborators contributed with collection of DNAs¹, phenotype identification of patients¹, BAL extraction², *BTNL2* cDNA cloning^{3,4} and expression experiments^{3,4}. I am very grateful to all collaborators for their contribution to the project.

In a candidate-gene association study, I have tested *CARD15* as sarcoidosis susceptibility gene.

No part of this thesis has been submitted to any other board for another qualification. Most of the results have already been published (see next page).

Kiel,.....(Ruta Valentonyte)

PUBLICATIONS RELATED TO THIS THESIS

1. Schürmann M, **Valentonyte R**, Hampe J, Müller-Quernheim J, Schwinger E, Schreiber S. *CARD15 gene mutations in sarcoidosis*. Eur Respir J, 2003. **22**(5): p. 748-54.
2. Hampe J, **Valentonyte R**, Manaster C, Teuber M, Jenish S, Entz P, Nagy M, Schreiber S. *Nonelectrophoretic method for high-throughput HLA-DRB1 group genotyping*. Biotechniques, 2004. **36**(1): p. 148-51.

CONGRESS ABSTRACTS

1. Schurmann M, **Valentonyte R**, Albrecht M, Hampe J, Müller-Quernheim J, Schwinger E, Schreiber S. *Genes of the CARD15/TLR4/NF-kB signalling pathway in sarcoidosis*. in WASOG. 2002. Stockholm, Sweden.
2. Schurmann M, **Valentonyte R**, Albrecht M, Hampe J, Müller-Quernheim J, Schwinger E, Schreiber S. *Study of Toll-Like Receptor Genes in Sarcoidosis*. in NGFN/DHGP Symposium. 2002. Leipzig, Germany.

12. Acknowledgements

I deeply appreciate the expertise and advices I received from Prof. Dr. Stefan Schreiber who provided me with the opportunity to work in the laboratory under his supervision. His concept of application of genetic epidemiology, excellent working conditions, large resources of laboratory materials and modern equipment were fundamental for success of this project. I want to thank to the supervisor of my PhD studies Prof. Dr. Silvia Bulfone-Paus from Research center in Bortel.. My special thanks go to Dr. Jochen Hampe, who created the laboratory platform structure, developed internal database system, statistical analysis software and provided the scientific basis of this thesis. His enthusiasm, new ideas, advices and motivation helped me much doing this work.

I thank the head of the Department of General Internal Medicine of the University Clinic Schleswig-Holstein Campus Kiel, Prof. Dr. Ulrich R. Fölsch for the opportunity to work in his hospital.

I am very grateful to Dr. Manfred Schürmann and Prof. Dr. J. Müller-Quernheim, our collaborators from the Institute of Human Genetics, University-Clinic Schleswig-Holstein, Campus Lübeck, and the Department of Pulmonary Medicine, University Hospital at Freiburg, for the entire sample of sarcoidosis patients' DNA. This study would not have been possible without the support of DNA and patients data. I thank them for the large contribution to the study providing sarcoidosis linkage results on chromosome 6 and characterising patient phenotype. Expert assistance of Melanie Albrecht preparing sarcoidosis sample DNA is much appreciated. I thank Dr. Karoline Gaede who provided the bronchoalveolar lavage samples.

Many thanks to Prof. Dr. Michael Krawczak for the help analysing and interpretation the peri-HLA region genotyping results. I also thank for the „Genetic epidemiology and population genetics’’ and „Bioinformatics’’ courses, which supported to a general genetic knowledge writing this thesis.

No thesis can be finalized without being read and approved by a thesis committee. I want to thank my thesis committee members for their guidance and assistance.

I am very thankful to Dr. Klaus Huse for the BTNL2 expression, splice variant definition and cloning experiments, which helped to define the right splice variants of BTNL2 and interpretate the association results.

I thank Prof. Monika Stoll and Dr. Almut Nebel for their experienced leading of the DNA plating laboratory, which was responsible for the sarcoidosis sample plating. Andreas Koch is appreciated much for the leading of genotyping platform that produced high quality data.

I am especially grateful to Dr. Annette Stenzel, who is a great example of scientific work, experiment design and management of different stuff. I am very grateful for the panel of diallelic SNP assays, designed for the disease-associated gene finding in the chromosome 6 peri-HLA region, introduction to a general statistical analysis and the first knowledge in the experiment management. I am also thankful her for regular messages about the boiling water...

I am appreciated to Dr. Philip Rosenstiel for the BTNL2 cloning experiments. I am also thankful for the expression experiments on cell lines and many advices concerning the cDNA cloning work.

I wish to thank the scientists of our research group Drs. Peter Croucher, Silvia Macheretti-Croucher, Georg Wätzig and Christine Costello for helpful discussions concerning biological and technical aspects of the study. I am very grateful to Peter who spend much time correcting my English language in this thesis and for advices concerning statistical analysis of genotyping data. In addition, Dr. Christine Costello is much appreciated for the social evenings, making the studies and living in abroad pleasant.

I thank to the colleagues from my „Main office“ team for a pleasant working and social atmosphere during the writing of thesis: Andre Franke, Weiyue Zheng and Janina Heinze. In addition, I am very thankful to Andre who helped me to translate the „Summary“ chapter into German language.

Many thanks go to my colleagues Carl Manaster, Tim Lu, Rainer Vogler and Marcus Teuber, who created our internal databases and analysis software used in this thesis. Marcus Will is appreciated for laboratory network system administration.

I am very grateful to laboratory technicians for high quality expert assistance. Many thanks to Tanja Wesse and Birthe Petersen for the genotyping of many SNPs, Tanja Engler and Tanja Henke for preparing DNA samples, Tam Wendt for sequencing support, Hebke Hinz and Nadine Teepe for DRB1 group typing and measurements.

I am especially grateful to my parents and brother, who supported me with the trust and love threw the last years of my life in Germany. Nuoširdžiausiai dekoju savo teveliams ir broliui už suteikta didžiule moraline parama, gyvenant toli nuo gimtųjų namu. Aciu už meile, šiluma, tikejima, stiprybe ir viska, kas mane ikvepe ir suteike man jegu atlikti ši darba.

In the past three years, I have met many nice people here in Germany. Their kindness has had a positive influence on my general well –behaviour, which has contributed to the completion of this thesis. I am grateful to my ex-colleague and very good friend Olga for a very good time spent with her. My last special thanks go to my friend Marcin, who supporte d me with the sense of joy all the time spent here in Kiel.