Development of an SNP map in the peri-MHC region on the human chromosome 6 as a tool to identify candidate genes for inflammatory bowel disease

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

vorgelegt von

Dipl.-Biol. Annette Stenzel

Kiel
2003

Referent:	Prof. Dr. Hans-Dieter Flad
1. Koreferent:	Prof. Dr. Stefan Schreiber
2. Koreferent:	Prof. Dr. Thomas Bosch

Tag der mündlichen Prüfung: 11.07.2003

Zum Druck genehmigt: Kiel,.....16.07.2003

gez. Prof. Dr. W. Depmeier (Dekan)

Table of contents

Abbreviations and symbols

AA	amino acid
ASP	affected sibling pair
bp	base pairs
BRPF3	bromodomain and PHD finger containing, 3
CARD15	caspase recruitment domain family, member 15 (synonym: NOD2)
CCD	charge coupled device (camera)
CD	Crohn's disease
cDNA	complementary DNA
CED	chronisch-entzündliche Darmerkrankungen
CEPH	Centre d'Etude du Polymorphisme Humain (cell lines)
cM	centiMorgan
CU	Colitis ulcerosa
°C	degree Celsius
χ^2	chi-square, measure of association or independence
D'	D prime (measure of LD)
dATP	2'-deoxyadenosine-5'-triphosphate
dCTP	2'-deoxycytidine-5'-triphosphate
DDW	double-distilled water
dGTP	2'-deoxyguanosine-5'-triphosphate
DNA	deoxyribonucleic acid
dNTP	2'-deoxynucleoside-5'-triphosphate
dsDNA	double-stranded deoxyribonucleic acid
dTTP	2'-deoxythymidine 5'-triphosphate
EDTA	ethylenediaminetetraacetic acid
FAM	6-carbofluorescein
Fig.	figure
Figs	figures
g	gram(s)
x*g*	relative centrifugal force (RCF)
h	hour(s)
HFE	hemochromatosis gene
HLA	human leukocyte antigen
HSP	heat shock protein (e.g. HSP-70)
HWE	Hardy-Weinberg-Equilibrium
IBD	Inflammatory bowel disease
ibd	identity by descent
ibs	identity by state
IFN	interferon (e.g. IFN-g)
IkB	inhibitor of NF-kB (e.g. IkB-a)
IL	interleucin

kb	kilobase
kD	kilodalton
l	liter(s)
LD	linkage disequilibrium
LDU	linkage disequilibrium units
LOD	logarithm of odds
Log10	decadic logarithm
LTA	lymphotoxin-alpha (synonym: tumour necrosis factor beta, TNF-β)
M	molar (mol/l)
MAPK	mitogen-activated protein kinase
Mb	Mega base
MC	Morbus Crohn
mg	milligram(s)
MHC	Major Histocompatibility Complex
min	minute(s)
MKK	mitogen-activated protein kinase kinase
ml	millilitre(s)
MLE	maximum likelihood estimate
mM	millimolar (mmol/l)
MRC	Molecular Research Center
mRNA	messenger RNA
Mxi2	Max-interacting protein-2
µg	microgram(s)
µl	microlitre(s)
µM	micromolar (µmol/l)
ng	nanogram(s)
nM	nanoMolar (nmol/l)
NOD2	nucleotide oligomerization domain-2 (synonym: CARD15)
NPL	nonparametric linkage
OR	odds ratio
p	probability
p. a.	*pro analysi*
p-ANCA	perinuclear antineutrophil cytoplasmic antibody
PCR	polymerase chain reaction
pH	potentia hydrogenii (hydrogen ion concentration)
pmol	picomol
r.t.	room temperature (ca. 20°C)
rmax	maximum radius (centrifuge parameter)
Rn	fluorescent emission of the normalised reporter dye
RNA	ribonucleic acid
rpm	rotations per minute (centrifuge parameter)
s	second(s)

SNP	single nucleotide polymorphism
Taq-polymerase	*Thermophilus aquaticus* DNA polymerase
TAMRA	6-carboxytetramethylrhodamine
TBE	Tris borate EDTA
TCR	T-cell receptor
TDT	Transmission disequilibrium test
TE	Tris EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
TET	tetrachloro-6-carbofluorescein
T_H	T-helper cell (T_H1 and T_H2: type 1 and 2, respectively)
Tm	melting temperature
TNFA	tumour necrosis factor alpha
TREM1	triggering receptor expressed on myeloid cells 1
Tris	tris-(hydroxymethyl)-aminomethane
UC	ulcerative colitis
UV	ultraviolet (light)
VIC	trade name for fluorescent dye

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 base nomenclature

A	Adenine
G	Guanine
C	Cytosine
T	Thymine

1. Introduction

1.1. The human peri-MHC region

1.1.1. General features of the human peri-MHC region

The human major histocompatibility complex (MHC) is an about 3.6 Megabases (Mb) large genomic region on the short arm of human chromosome 6 (1, 2). The area is subdivided into three parts called class I, II and III according to the function and protein structure of the genes first located there. While class I is most telomeric on the chromosome 6p and class II most centromeric, the class III is centric to the other two. From the 224 gene loci identified in the human MHC 128 are predicted to be expressed, and it is estimated that about 40% of these are functional in the immune system, while many are still of unknown function (3). The average gene density (including pseudogenes) over the entire MHC is one gene per 16 kilobases (kb) with distinct regional variations. While the class I and II have a high rate of pseudogenes, the class III seems to contain only few pseudogenes (3). Within the class I and II gene cluster the most polymorphic human proteins have been found, some of which have over 200 allelic variants (3). The classical HLA antigens encoded in each region are HLA-A, -B, and -C in the class I region, and HLA-DR, -DQ and -DP in the class II region (Fig 1.1). All class I genes are between 3 and 6 kb length, whereas class II genes are 4-11 kb long (4). Analysis of the immediate flanking regions revealed that the 'classical' class I and class II regions extend further than previously thought (5, 6). The 1998 Nomenclature Committee recognized additional HLA genes all of which are in the class I and Ib regions: HLA-E, -F, -G, -H, -J, -K and -L (7). Among those, only HLA-E, -F and -G are expressed (8) Besides the classic HLA genes the region contains a number of related genes that are not accounted to the classic HLA genes (MICA etc.).

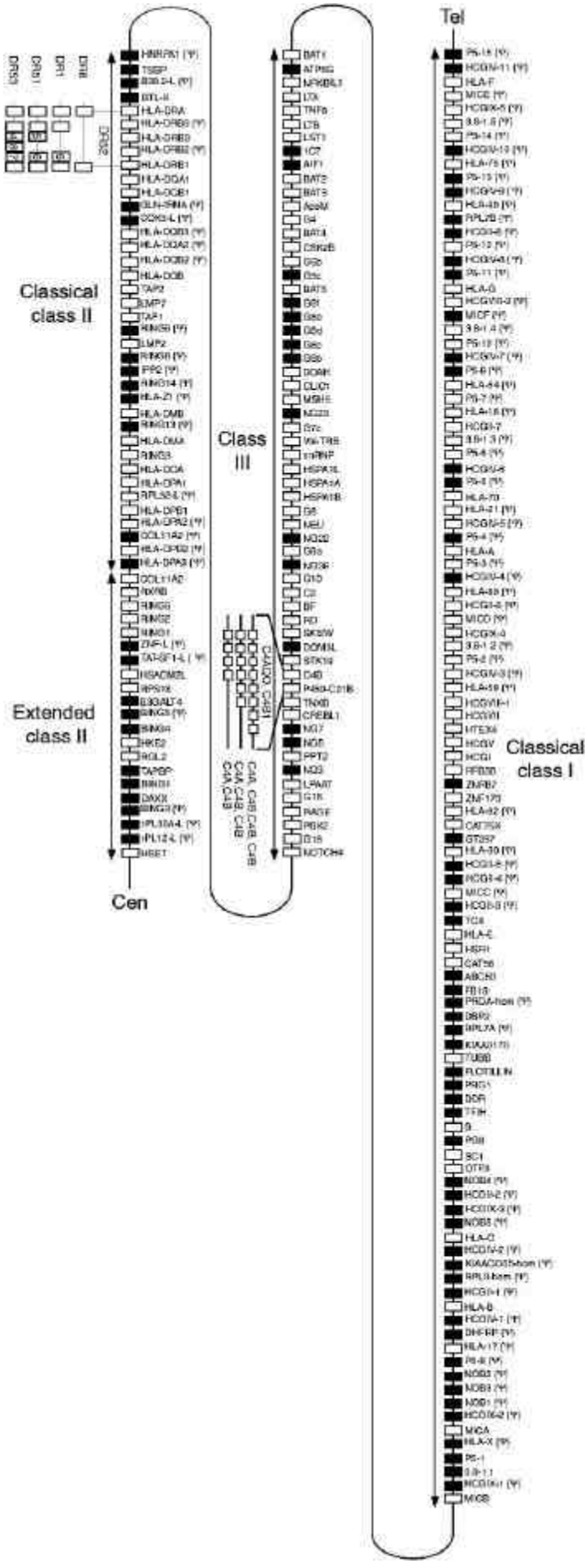

Fig. 1.1: Complete gene map of the human MHC as published by The MHC sequencing consortium [consortium, 1999 #49].
Genes are in the order from telomere to centromere but not to scale. Indicated by filled boxes are gene loci that were discovered or located to the MHC as a direct result of the genomic sequence. The other genes had been located in the region before. In regions where different haplotypes are known they are indicated.

1.1.2. Function of genes in the human MHC

The products of genes within MHC play a fundamental role in the regulation of immune response. The HLA class I and II genes encode for cell surface glycoproteins that are relevant in the adaptive immune response to present antigens to T-cells. The antigens, small protein fragments (peptides of 7-12 amino acids) of pathogens, are fixed in a binding cleft of the HLA glycoproteins and are carried to the cell surface where they can be detected by T-cells. The 2 classes of HLA molecules are specialized to present different sources of antigens. HLA class I molecules present endogenously synthesised antigens, e.g. viral proteins. HLA class II molecules present exogenously derived proteins, e.g. bacterial products or viral capsid proteins. Class I molecules can be found on each body cell. Antigens presented through class I molecules are recognized by T-cell receptors (TCR) of cytotoxic T-cells that express the same HLA antigen, resulting in the destruction of the infested cell. HLA class II surface glycoproteins are usually expressed only on activated leucocytes, the expression can be activated in certain cells that normally do not express class II molecules by stimulation with Interferon-γ (IF-γ). T-helper cells can recognize antigens presented this way.

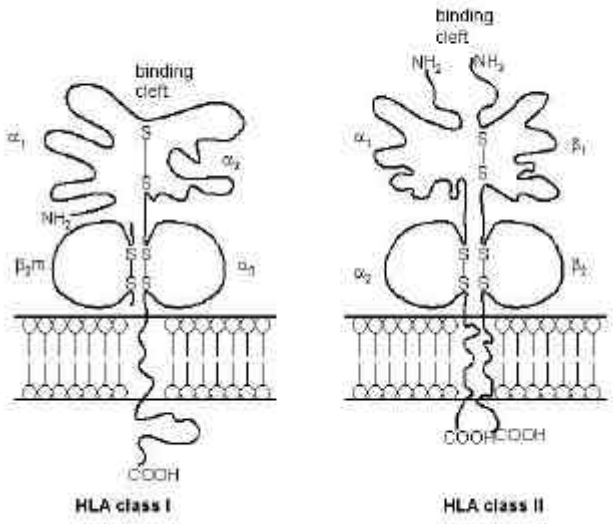

Fig. 1.2: Schematic representation of HLA class I and II cell surface glycoproteins.
The epitopes in the α-1 and α-2 domains of the class I and α-1 and β-1 of the class II molecule are specific for each individual and are themselves an antigen target in the immune response to transplanted tissue.

Class I and II molecules are transmembrane heterodimer from two different polypeptide chains (α– and β-chains) that are not covalently bound (Fig. 1.2). For the class I molecules only the variable 45 kilodalton (kD) α-chain is encoded in the MHC region, the β-chain is the 12 kD β_2-microglobulin (encoded on the human chromosome 15q). The α-chain is divided in

3 extracellular globular domains, the transmembrane region and a cytoplasmatic tail. The variable antigen-binding cleft is formed by the highly variable α_1 and the α_2 domain and can contain polypeptide chains with up to 12 amino acids. The β_2-microglobulin is attached by hydrophobic interaction and has no transmembrane compound. For the class II molecule both chains are encoded in the MHC region (A and B), and both contain 2 extracelluar globular domains (α_1, α_2 and β_1, β_2) a transmembrane region and an intracellular anchor domain. The α-chain can vary between 30 and 34 kD, the β-chain between 26 and 29 kD. The antigen-binding cleft is formed by the highly variable α_1 and β_1 domains of the two chains and by this has no direct limitation to the length of the polypeptide presented. While the β_2-microglobulin and the α_3 domain from HLA class I and the α_2 and β_2 domains from HLA class II show high similarity to immunoglobulins, the other domains do not exhibit any similarity. A highly relevant feature of the HLA molecules is their co-dominant expression, both alleles contribute equally to the phenotype. Together with the diversity in the HLA molecules, this generates a great variety of antigen binding clefts, which means a wider potential immune response to pathogens.

In addition to the highly polymorphic HLA class I and II genes, several other immune-regulatory genes are known to be located in the MHC region (3). The Class III region is not actually a part of the MHC complex, but is located within the MHC region, and its components are either related to the functions of HLA-antigens or are under similar control mechanisms to the HLA genes. Genes of class III region encode structurally and functionally diverse molecules such as serum complement proteins (C2, C4A, C4B, BF), which are part of the innate immune response (9, 10), tumor necrosis factors (TNFA, LTA) which are cytokines essential in the regulation of the immune response, heat shock proteins (HSP) and adrenal 21 hydroxylase enzyme (CYP21B) which is important in the corticosteroid metabolism (4, 11, 12).

In the extension of the HLA region centromeric to class II there is a significant number of genes involved in the immune regulation pathways such as MAPK14, FKBP5, the binding of the HLA molecules (TAPBP) and genes related to apoptosis (DAXX). Furthermore the HFE gene, important in hemochromatosis and in structure similar to the HLA class I genes and some of the genes encoding for histone proteins are positioned to the telomeric side of the HLA class I (6).

1.1.3. Polymorphism and genomic organization of the human peri-MHC region

The extreme polymorphism is one of the most prominent features of the MHC. Variation levels of 5-17% have been reported at some loci (HLA-DP, DQ; B and C), which are the highest levels found in the human genome so far (3, 13-16). While the majority of variation results from polymorphisms in the exons, the classical HLA-like genes have nearly identical intron/exon structures in all vertebrates and classical MHC molecules can be found throughout the vertebrates (17-20). Most affected by the high degree of variation are those exons coding for the antigen binding regions providing a means against the large variety of pathogens. Important is the selection of T-cells during their thymic development to ensure the deletion of variants with a high affinity to "self" peptide complexes. This is a requirement to prevent immune response to body innate peptides and MHC molecules. Through this step the immune response phenotype of the individual is determined. The ability to present antigens is dependent on the variation and therewith the amino acids present in the binding cleft. Some alleles might be more efficient, others less, in binding an antigen, which has consequences in the ability to trigger the immune response (21).

Despite the enormous number of alleles at each expressed loci, the genes of the MHC region are normally inherited unchanged from the parents, so that one set of allelic variants of the HLA genes present in the parents is present in the same form in the offspring. This phenomenon caused the formation of so called haplotypes where certain alleles of different gene loci tend to segregate together rather than randomly. The effect is caused through linkage disequilibrium (LD) between gene loci, which is described as decreasing with increasing physical distance. The phenomenon of LD will be addressed later in more detail. The number of haplotypes observed in populations is much smaller than theoretical expectations. The formation of haplotypes can be observed to different degrees between the HLA genes, as in other areas of the genome. At HLA class II, this phenomenon is so pronounced, that the presence of specific HLA-DR alleles can be used to predict the HLA-DQ allele with a high degree of accuracy before testing (22). The assignment of association signals of diseases to HLA class I or II genes or adjacent non-HLA genes is therefore difficult in the MHC region. Haplotype structures within the MHC region have been characterized through classical HLA typing, and high levels of LD have been detected for HLA class I and II genes, between certain class I and II genes, and with neighbouring loci in the class III region (23, 24). An average distance between adjacent recombination hot-spots of 0.8 Mb has

been demonstrated in single sperm analyses, but these data also showed a high degree of variation. Between hot-spots, comparatively stable blocks comprising some 100 kb of low level recombination occurred, which were separated by less active spots of recombination ("warm spots")(25). It appears likely that the mammalian MHC genes have developed their enormous diversity through a combination of frequent recombination and diversifying selection (26). This implies that neither the concept of rapid LD decay with increasing distance nor that of a stringent organization into well-defined haplotype blocks may be fully compatible with the LD pattern pertaining to this region.

1.1.4. MHC and disease

The importance of HLA genes in antigen presentation and regulation of immune response results in association of the region with more diseases than any other region of the human genome. The MHC region therefore represents a primary target for disease gene discovery efforts. Most of these diseases are of chronic inflammatory or autoimmune background. For some diseases a linkage with the HLA region has been established as for example rheumatoid arthritis (27), psoriasis (28), sarcoidosis (29) and chronic inflammatory bowel disease (IBD) (30). In many cases the exact gene is not identified but associations to a number of HLA genes have been established. For many more diseases association to the HLA region was shown as for example diabetes mellitus (31, 32), and multiple sclerosis (33). A great number of studies have analysed potential association with certain HLA gene alleles or haplotypes, often with contradictory results. In a few cases a disease gene with causative polymorphisms was successfully identified, as in the case of hemochromatosis (HFE) (34).

1.2. Methods of disease gene detection

1.2.1. Genome mapping linkage analysis

Once epidemiological evidence shows a genetic background for a disease, the next important step is to identify the gene or, for polygenic diseases the genes, that carry the disease causing mutation. Considering that 20,000 genes are presently known and the amount of genes that is estimated for the human genome, estimations range between 28,000 and 120,000 (35), most estimates range between 30,000 to 50,000 genes (36, 37), it is necessary to reduce the area of the genome in which such a gene could be situated. One method developed for that purpose is the genome-wide linkage analysis. The unknown position of the potential disease gene is thereby estimated similar to the establishment of a genetic map on the basis of recombination frequency. The method employs polymorphic markers of which the genetic and physical positions are known or are estimated as well on the basis of the recombination fraction, which can vary from 0 (no recombination, the two markers are completely linked) to 0.5 (no linkage between the markers). The disease itself is handled like a polymorphic marker with 2 variants, affected and unaffected for a qualitative trait, but of unknown position. For quantitatively measurable traits the measured values are used. Recombination between each of the genetic markers and the disease "marker" is determined based on the frequency each allele was observed. The maximum likelihood that the observed data are caused by a determined recombinant fraction is tested. The likelihood for each value of the recombinant fraction between pairs of markers is compared and thereby the odds ratio and the logarithm of the odds ratio (LOD score) are determined. The LOD score indicates the likelihood of linkage (38, 39). This is the logarithm of 10 from the probability that two markers are linked with a given recombination value divided by the probability that they are unlinked. The most probable position of the disease "marker" is between those 2 genetic markers where the smallest recombination frequency is measured, indicating the strongest linkage and the shortest genetic distance. The maximum likelihood estimate (MLE) is the value that gives rise to the largest value of the likelihood (or LOD score) (40). The MLE is highly efficient, in the sense of the precision obtained using a given set of data. With increasing sample size it will converge closer to the true value. For small samples, the MLE value might tend to be lower or higher than the true value. The steepness of the LOD curve around the MLE reflects the precision of the estimate (40). In the analysis of a qualitative trait usually the nonparametric linkage (NPL) is measured, where the algorithm is adapted to the bipolarity of the trait. The

usual method is the multipoint linkage analysis where several markers at the same time are used to map the disease locus. For an initial screening a density of one marker every 10 cM is aimed at. The markers employed for this are microsatellites. They are abundant in the human genome so that an even spacing can be achieved easily and due to their great number of variants they are of high information content. The analysis of the polymorphic markers is performed in affected sibling pairs (ASP). They share the disease allele through identity by descent (ibd). With the determination of the allele status at the polymorphic marker recombination can be detected between the marker and the disease allele. The closer the disease allele is linked to the marker, the less recombination can be expected and the higher is the resulting LOD score. If one or more regions in the genome scan are marked with high LOD scores, a fine mapping of that region can follow. In that situation a density of one marker per 1 cM is aimed at. Taking the possibility of false positive results in account, a critical value of $\geq$ 3.6 for the LOD score was suggested (41). However LOD scores depend on the density and information content of the marker, the size of the ASP population and preciseness of the disease phenotype.

1.2.2. Association mapping analysis

Association analysis is useful to refine the location of major genes prior to positional cloning or the candidate gene analysis. A linkage region even after fine mapping might be several Mb long and contain several hundred genes. For this analysis the density of markers is increased (3 to 50 kb) and usually diallelic single nucleotide polymorphisms (SNPs) are employed as markers. Association studies can be family or population based resulting in two different analysis methods, the transmission disequilibrium test (TDT) for family based studies (42) and the case control analysis for population based studies.

In a population based study with the case control set-up, the Pearson's χ^2 statistic (or any other appropriate χ^2 statistic) is calculated. Thereby the number of each allele or genotype present in each disease category (case and control) is transferred into a contingency table (43). The significance at the 95% confidence interval is determined which is depending on the degrees of freedom (df) in the contingency table (1 for allele association and 3 for genotype association). Additionally the odds ration or the relative risk can be calculated for interesting markers.

For the TDT study design the marker analysis involves families with at least one affected offspring (monoplex families) but may have more (multiplex families) and the healthy parents (44). The transmission of alleles from a heterozygous parent to an affected offspring is tested and compared to the untransmitted alleles. Different methods to evaluate the transmission to the affected offspring have been developed. Either transmitted alleles are compared direct to untransmitted alleles or the difference to the number of alleles expected to be transmitted according to HWE is analysed (45-47). TDT association studies are considered to be more sensitive (having greater power) than linkage methods to detect genes with small to moderate effects on disease (47). Compared to the population based association analysis the TDT is not affected by stratification in the population.

1.2.3. Candidate gene analysis

Candidate genes are characterised through their possible functionality within a disease or trait phenotype. They might be identified by their structure or their role within a biochemical pathway (biological candidates). Genes that have been identified for similar traits in other species or animal models for a disease phenotype might help to identify a homologue gene in the species of interest (comparative candidates). Once a genomic region has been identified through linkage or association analysis the search for candidate genes can focus on that region (positional candidates). The procedure of analysis in a candidate gene approach is to perform sequence analysis for the detection of mutations, with emphasis on the coding sequence. This could be a SNP, an insertion or a deletion. Depending on the location of a mutation this could have different impacts on the functionality of the encoded protein. SNPs in exons could lead to an exchange in the amino acid (AA) chain, or the truncation of the protein if the new base causes a stop codon. Insertion or deletion mutations could result in a shift in the reading frame leading to a completely different protein. SNPs in the introns could in some cases change the affinity to enhancer or silencer proteins and a mutation in the promotor region could alter the level of expression of the protein. Such a variation is then tested in an association study. A positive association can either identify the causative mutation itself or is linked to another variation on the same haplotype. This should be confirmed through an analysis of the biological relevance of the polymorphism. A direct functional impact of an SNP in the coding sequence causing a change in the amino acid sequence, with consequent alteration of the phenotype is the most probable situation. The proof of functional relevance of polymorphisms in the promotor region or in introns is somewhat more complicated.

1.2.4. The problem of genome haplotype structure

For the understanding of the genetic determination of human phenotypic characteristics, including the predisposition to disease, the characterization of human genetic variability and its correct interpretation are important. Therefore, considerable research efforts have been undertaken in order to quantify the statistical association between physically linked genetic variants in the human genome, and to classify the concomitant haplotype diversity. The amount of haplotype diversity in a population is inversely related to the level of linkage disequilibrium (LD) between neighbouring loci, i.e. to the excess in co-occurrence in germ cells of particular allelic variants of these loci. The usually observed decrease of LD between loci of increasing physical and genetic distance broadly reflects their gradually increasing likelihood of recombination (48). This inverse relationship between LD and distance should - at least in principle - facilitate the positional cloning of disease-associated genetic variants through the analysis of surrogate markers located close to the respective susceptibility genes. Under a model of nearly neutral evolution, any newly introduced mutation would initially be of low population frequency but at the same time would show strong LD with the alleles defining the haplotype background on which it first arose. Over time, the population frequency of the mutation must increase in order for it to be detectable in a given extant population, but LD would decrease during the same process due to recombination (49, 50). Recent empirical evidence suggests that the relationship between LD and distance is not uniform. Instead, the human genome appears to be organized into regional blocks with high LD within, and little or no LD between blocks (51). According to this supposition, recombination events are located at so-called “hot-spots" (52) that define intermittent blocks, each represented even in different populations by a few common haplotypes. Under such conditions the construction of systematic haplotype maps would be possible and useful (53, 54). Whilst the existence of haplotype blocks and recombination hot-spots has been convincingly demonstrated, the criteria for the definition and localization of haplotype blocks are still subject to discussion (55, 56). The design and efficiency of association gene mapping studies is critically dependent upon a detailed knowledge of the local LD pattern.

1.3. Inflammatory bowel diseases (IBD)

1.3.1. Pathogenesis and pathophysiology of IBD

Inflammatory bowel disease (IBD) refers to a complex chronically relapsing autoimmune disorder of the gastrointestinal tract of unknown etiology. It has been classified into Crohn's disease (CD) and ulcerative colitis (UC) based on clinical, radiologic, endoscopic and pathological criteria (57). Active inflammation is marked by chronic diarrhea, rectal bleeding, abdominal pain, fever, joint pain, and weight loss. These symptoms can range from mild to severe, and may gradually develop from an initial minor discomfort, or may present themselves suddenly with acute intensity. Between periods of active inflammation, the disease can be inactive with a low degree or no inflammation (remission) for days up to years. During remission, patients suffer few, mild or no symptoms. The disease incidence peaks in early adulthood and is rising worldwide, with a current lifetime prevalence of 0.1–0.5% in Western countries (58, 59). IBD is a multifactorial disease caused by the interplay of genetic, environmental and immunological factors (60). Epidemiological and genetic linkage studies in IBD provide a thorough proof for a genetic background (61, 62), and a first disease gene (NOD2 or CARD15) has been identified for CD. (63-65). No single genetic variant is sufficient to initiate the disease. In addition, complex environmental triggers are required for disease expression. The lifestyle of an industrialized society seems to be important for the penetrance of the genetic factors (66).

In CD inflammation can arise anywhere in the digestive tract from the mouth to the anus. The site of inflammation determines the range of symptoms. In the intestine CD manifests as patchy skip lesions (diseased segments of bowel interspersed with normal segments). Aggregates of immune cells consisting of T cells, monocytes and macrophages in the bowel wall can be observed in half of the cases. The structure of the mucosa is distorted and shows progressive atrophy. From aphthous and focal mucosal ulcers in early disease, linear ulcers can develop. CD affects all strata of the intestinal wall and other organs in contact with the intestine may become involved through transmural inflammation. This leads to the formation of fistulae, *e.g.* between the intestine and bladder or other inner organs. Inflammation, oedema or fibrotic scaring in the healing process can cause strictures of the bowel lumen.

Inflammation in UC starts in the rectum and extends proximally up the bowel. The damage of the mucosa is continuous from the rectum to the proximal colon without evidence of skip lesions. The inflammation is located predominantly within the lamina propria of the mucosa. Scattered crypt abscesses with ulceration and islands of regenerating mucosa forming pseudopolyps are observed. A 'backwash ileitis'(with a wide open ileocecal valve and dilated terminal ileum) may involve the ileum and appendix and can occur in 10% of the cases of serious pancolitis. A severe course of UC is the formation of a toxic megacolon, a sudden cessation of bowel function leading to toxic dilatation and eventual perforation of the bowel. As only the inner lining of the intestine is affected in UC, nearby organs are not affected by the formation of fistulae. Both CD and UC can lead to numerous extra-gastrointestinal inflammatory manifestations in the liver, eyes, skin and joints (57).

The gut lamina propria is constantly exposed to antigens and pathogens coming along with the nutrition. A low level of 'physiological' intestinal inflammation is normally maintained by the immune system to prevent infection. The enteric flora and food antigens are thought to induce and sustain this state. In IBD, this carefully balanced state is altered, which results in a destructive and persistent 'pathological' inflammation. After a primary inflammatory process microbes and microbial products may infiltrate the submucosa or lamina propria, resulting in a reinforced inflammation leading to an increased mucosal damage (67).

When the state of inflammation shifts from the physiologic to the pathogenic state the pro- and anti-inflammatory cytokines become imbalanced (68, 69). In CD a predominant T_H1 response with elevated levels of IL-2 and IFN-γ was observed in studies (70, 71). In UC humoral immunity with an increased level of IL-5 and IL-10 seems to predominate, but evidence for a classical T_H2 situation is scarce (60, 72). In both CD and UC, activated macrophages participate in the mucosal immune response, *e.g.* by producing pro-inflammatory cytokines such as TNF-α, IL-1β and the chemokine IL-8 (73-75). It has been suggested that TNF-α plays a central role in the pathogenesis of CD and is likely to be at the apex of the inflammatory cascade (69, 71, 73, 76).

In almost half of UC affected individuals and in a small group of CD cases a perinuclear antineutrophil cytoplasmic antibody (p-ANCA) can be detected. It seems to be more prevalent in more aggressive UC. P-ANCA reactivity is suggested to derive from the recognition of heterogeneous neutrophil-associated antigens (77). A sub group of the antibody, called

atypical p-ANCA, recognizes a 50 kD myeloid-specific nuclear envelope protein that can be detected in UC affected carriers of the antibodies (78).

1.3.2. Genetic background of IBD

Epidemiological investigations of IBD have consistently shown familial clustering with a 10 fold increase in the familial risk of UC or CD (79). This and an increased concordance of the IBD phenotype in monozygotic twins (80, 81) clearly show a genetic cause of the disease. In 25% of multiply affected families cases of CD and UC are present (82), indicating that at least some of the risk alleles will be common to UC and CD. Genetic linkage studies in IBD affected sibling pairs support this with susceptibility loci on the chromosome 1 (83), chromosome 5 (84, 85) called IBD5, chromosome 6 (30, 85) named IBD3, chromosome 12 (86) identified as IBD2, chromosome 14 (87, 88) called IBD4, chromosome 16 (86, 89) IBD1, the first IBD locus determined, and chromosome 19 (85), and as well on chromosome X (90).

Within the linkage region on chromosome 16 (IBD1) a first disease gene CARD15 (or NOD2) has been identified. Variants of the NOD2 gene (C14772T (R702W) in exon 4, G25386C (G908R) in exon 8 and 32629insC (1007insC)) are highly associated with CD (63-65), but cannot account for all cases. NOD2 encodes a protein with homology to disease resistance gene products in plants. Among other functions, NOD2 may be involved in NF-κB signalling (67). Besides this first gene, the multiple linkage regions and segregation models indicate that more than one risk allele is involved in the pathogenesis of IBD (91, 92).

1.3.3. IBD and the human MHC

An IBD susceptibility gene in the MHC region on chromosome 6 (IBD3) has been suggested through linkage and association analysis methods. An initial linkage analysis in our group employing non-parametric linkage analysis on 353 sibpairs with 17 markers on chromosome 6 (average spacing 9.4 cM) showed a LOD score of 2.1 for IBD (93). As in other screenings the linkage on chromosome 6 did not achieve the proposed significant LOD score of $\geq$ 3.6 (86, 94). Supportive evidence for a linkage on chromosome 6p was given by a multiple regression analysis in 49 families with CD (95). In order to support or reject the linkage to IBD the marker density in the region on chromosome 6p including the MHC (microsatellite

markers: D6S309 - D6S257) was increased by 11 microsatellite markers and the study population was expanded to 428 ASPs total. Additionally all five known functional single nucleotide polymorphisms in the TNFA and LTA genes were tested (30). In this analysis we observed a peak LOD score of 4.1 in IBD in the HLA region (microsatellite markers: D6S470–D6S271) (30). Linkage was stronger in CD but present for UC as well (Fig. 1.3). Evidence for linkage in the MHC region has been repeated in other samples (85, 96).

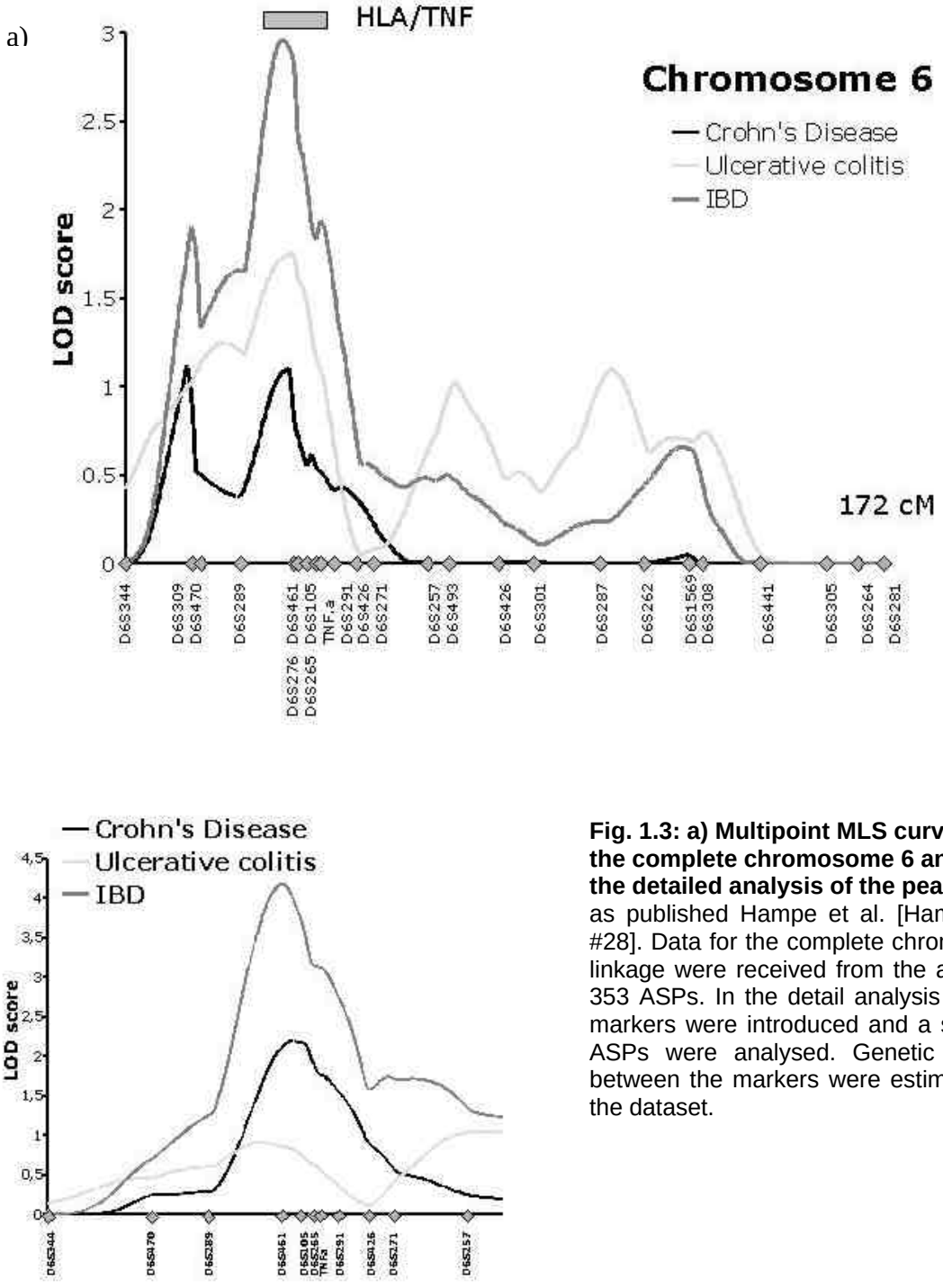

Fig. 1.3: a) Multipoint MLS curves for the complete chromosome 6 and b) for the detailed analysis of the peak region, as published Hampe et al. [Hampe, 1999 #28]. Data for the complete chromosome 6 linkage were received from the analysis of 353 ASPs. In the detail analysis additional markers were introduced and a set of 428 ASPs were analysed. Genetic distances between the markers were estimated from the dataset.

In the past 25 years more than 50 studies evaluating IBD association with polymorphisms in a limited number of genes in the MHC have been published. Several studies show positive associations with class I allele variants or haplotypes (97-100). A number of studies have described association to HLA class II gene variants or haplotypes (101-104). Genes from the HLA class III like the immune regulatory important TNF-α have been under investigation as well (105, 106). However other publications decline associations to genes in the MHC described (30, 107-110). Some studies show that associations of HLA genes with UC (111, 112) differ from associations with CD (113, 114). A meta-analysis of a great number of association studies indicated that DR2, DR9, and DRB1*0103 are positively associated with UC, a negative association was found for DR4 and UC. For CD a positive association was found with DR7, DRB3*0301, and DQ4 and a negative association with DR2 and DR3 (115).

Within the MHC region, the identification of a causative mutation is complicated through the high density of genes and the extent of LD. The polymorphism in the HLA genes and the great number of possible haplotypes as well as the variation of more frequent haplotypes between different populations reflect the limitations of association studies in the MHC region. Despite the ambiguity of the study outcomes, the HLA class I and II genes remain important to be studied, due to their antigen presenting function.

TNF-α is a key inflammatory mediator in the pathophysiology of IBD (69, 71, 73, 76). An association of an inferred haplotype (TNFa2b1c2d4e1) of microsatellites flanking the TNF locus with CD has been described (105), yielding a p-value of 0.01 (odds ration 4.4). In a direct investigation of polymorphisms of the TNFA gene promoter the uncommon allele 2 of the –308 TNFA promoter polymorphism was found to be decreased in UC (p=0.044) (106). Linkage and association analysis of 3 TNFA promotor polymorphisms and 2 LTA polymorphism declined a direct connection between TNF and IBD (30). Even though the TNF-α by itself might not be the susceptibility gene for IBD in the chromosome 6 linkage region, genes coding for proteins involved in the downstream signalling pathway might be directly associated and the association observed to alleles of the TNFA gene might be caused by linkage to other variants in nearby inflammatory relevant genes. One of these genes might be a mitogen activated protein kinase (MAPK14 / MAPK13) that are located centromeric of the TNF. The activation of the MAPK14 protein upregulated the expression of TNF in vitro (116). The activity of the MAPK14 is upregulated in the inflamed mucosa of patients with both CD and UC resulting in a substantial increase of the active form of p38α (117).

Activation of MAPK14 and the isoforms results from phosphorylation in response to treatment with proinflammatory cytokines or exposure to environmental stress (118).

1.4. Aims of the study

A linkage region for IBD has been established and confirmed for the human chromosome 6p21. Through fine mapping linkage analysis with additional markers and ASPs the region was located between the microsatellite markers D6S461 and D6S271, which covered about 20 Mb including the peri-MHC region. In order to further confine the region where a potential gene with a disease causing mutation might be located and to eventually identify the gene carrying the causative mutation the aims of the present work were

(i) to establish a dense map with SNP markers and to analyse association with IBD within the region defined.

(ii) to identify functional and positional candidate genes from the association study and analyse for coding polymorphisms that could be causative for IBD.

(iii) to reveal underlying structure of linkage disequilibrium between markers that could account for association signals distorted from their origin.

2. Materials and methods

2.1. Materials

Table 2.1 Materials part 1

Material	Manufacturer / Supplier
100 bp 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
AmpliTaq® DNA Polymerase	Applied Biosystems; Weiterstadt, Germany
AmpliTaq® Gold DNA Polymerase	Applied Biosystems; Weiterstadt, Germany
AmpliTaq® DNA Polymerase	Applied Biosystems; Weiterstadt, Germany
Bacillol®	Bode Chemie; Hamburg, Germany
BigDye Terminator Ready reaction kit	Applied Biosystems; Weiterstadt, Germany
Bromophenol blue	Sigma; München, Germany
Cell culture flasks (250 ml; canted neck)	BD Biosciences; Heidelberg, Germany
Cryotubes (2ml)	Greiner Bio-One GmbH; Frickenhausen, Germany
DNAzol®	Molecular Research Center, inc. Cincinatti, Ohio, USA
dNTP set (100mM solutions 100µM each)	Amersham Biosciences; Freiburg, Germany
Easy peel heat seal foil	ABgene, Epsom, UK
EDTA	Sigma; München, Germany
EDTA blood vial 9 ml	Sarstedt; Nümbrecht, Germany
Ethanol *p. a.*	Merck; Darmstadt, Germany
Ethanol *technical*	Bundesmonopol für Branntwein (BfB); Offenbach, Germany
Ethidium bromide solution (10mg/ml)	Invitrogen; Karlsruhe, Germany
ExoI (Exonuclease I)	Amersham Biosciences; Freiburg, Germany
GeneAmp PCR buffer system (10x buffer w/o $MgCl_2$; 25mM $MgCl_2$ solution)	Applied Biosystems; Weiterstadt, Germany
Glycerol	Sigma; München, Germany
Invisorb Blood Giga Kit	Invitek, Berlin, Germany
isopropanol	Merck; Darmstadt, Germany
$MgCl_2$	Merck; Darmstadt, Germany
MicroAmp® optical 96 well reaction plate	Applied Biosystems; Weiterstadt, Germany
MicroAmp® single strips	Applied Biosystems; Weiterstadt, Germany
MicroAmp® single tubes	Applied Biosystems; Weiterstadt, Germany
Microtiter 384 well plates	Sarstedt; Nürnberg, Germany

Table 2.1 Materials part 2

Material	Manufacturer / Supplier
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; Nümbrecht, Germany
Microtiter strips (Nunc-Immuno Maxisorp™, 8-well, flat bottom) with 96-well frame	Nunc; Wiesbaden, Germany
Multiscreen column loader	Amersham Biosciences; Freiburg, Germany
PEQLAB DNA isolation system	PEQLAB Biotechnology GmbH; Erlangen, Germany
PicoGreen	Molecular Probes Europe BV, Leiden, The Netherlands
Pipette (serological, sterile with filter 5 / 10 / 25 ml)	Sarstedt; Nümbrecht, Germany
Pipette tips with filter (10 / 200 / 1,000 µl)	Sarstedt; Nürnberg, Germany
protein-kinase K	Invitek, Berlin, Germany
proteinase K	Molecular Research Center, inc. Cincinatti, Ohio, USA
saccharose	Merck; Darmstadt, Germany
SAP shrimp alkaline phosphatase	Amersham Biosciences; Freiburg, Germany
Sephadex powder (G50 superfine)	Amersham Biosciences; Freiburg, Germany
Sephadex 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; Weiterstadt, Germany
TBE buffer 10x ready pack	Amresco; Solon, OH, USA
TEMED	Sigma; München, Germany
Tris	Merck; Darmstadt, Germany
Triton-X	Sigma; München, Germany
Trypsin / EDTA (0.25% / 1 mM)	Invitrogen/Gibco; Karlsruhe, Germany
Tubes (0.5 / 1.5 / 2.0 mL)	Eppendorf; Köln, Germany
Tubes (0.5, 1.5, 2 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

2.2. Participants and study design

2.2.1. Association study population

In the association study with an SNP map of about 20 Mb including the peri-MHC region three groups of participants were involved. One group consisted of a large German family sample of IBD patients (affected sibling pairs, ASPs) and their healthy parents (multiplex families). This included 178 families with a total of 386 affected individuals, 254 with CD and 132 with UC. The second group contained 461 monoplex families with only one affected individual and both healthy parents sampled. In this group there were 306 CD and 155 UC affected individuals. In both cohorts the affected individuals and both healthy parents were sampled whenever possible although some few families remained incomplete with only one parent. Thirdly a group of 550 blood donors was collected as unrelated control individuals. For each affected individual the diagnosis of either CD or UC was confirmed by standard diagnostic criteria (57, 119).

Table 2.2: Population sample for the association study and the analysis of the candidate genes.

Study population	IBD cases	CD cases	UC cases	unaffected family members	number of families	unrelated healthy controls
monoplex families	461	306	155	812	461	548
multiplex families	386	254	132	274	178	
total	847	560	287	1086	639	548

Ascertainment criteria had to be determined prior to the initiation of patient collection. This included that clinical, radiological and endoscopic (type and distribution of lesions) examinations confirmed the diagnosis of CD or UC. Histology results needed to be consistent with the diagnosis. Other diseases, especially irritable bowel syndrome and infectious colitis, needed to be excluded firmly. In cases of uncertainty the patient was excluded from the study (30, 93, 120, 121). Patients and their family members were recruited at the 1st Department of General Internal Medicine at the University Clinic Schleswig-Holstein, Campus Kiel (Kiel, Germany), the Charité University Hospital (Berlin, Germany) and other collection centres in Germany as outlined in the appendix 8.3. The blood donors were collected through the Department of Transfusion Medicine at the University Clinic Schleswig-Holstein, Campus Kiel (Kiel, Germany). The detailed sample population overview is given in Table 2.2. The

same population was employed as well for the genotyping analysis of SNPs defined through sequencing in the candidate genes.

2.2.2. The population for HLA-DPA1 analysis

The European population included into this analysis was part of the original chromosome 6p linkage study. The German families coincided with the multiplex family population used for the association study. Additionally multiplex families from the UK (48% of the total population) and the Netherlands (6%) were included. The families from the UK were sampled through the King's College School of Medicine, Guy's Hospital, and the ST. Mark's Hospital London (UK) and the Samples from the Netherlands through the Academic Medical Center, Amsterdam (The Netherlands). A total of 249 families were included with 527 IBD affected individuals (292 with CD and 235 with UC). As unrelated healthy European controls 174 blood donors from the control group described above were comprised (Table 2.3). In the South-African population, families with one affected individual with the diagnosis CD or UC and one or more unaffected first-degree family member were sampled. These were either healthy parents or in some cases healthy siblings. A total of 50 families with 30 CD and 20 UC affected individuals and 82 related healthy controls participated. Twice only the patient was included. The families from South Africa were from a variable ethnical background, with 17 of white and 31 of mixed coloured origin. Collection of samples and ascertainment of the diagnosis according to the above mentioned criteria was performed at Groote Schuur Hospital, University of Cape Town, Department of Internal Medicine (Cape Town, South-Africa).

Table 2.3: Population sample for the HLA-DPA1 analysis.

Study population	CD cases	UC cases	IBD cases	Related healthy controls	number of families	unrelated healthy controls
European	292	235	527	380	249	174
South-African	30	20	50	82	48*	-
South-Korean	23	61	84	-	no families	71

* from two families only the patient

A population of 84 unrelated individuals diagnosed with CD (23) or UC (61) according to the inclusion criteria were recruited from referral patients at Yonsei University, College of Medicine (Seoul, South-Korea). Seventy-one unrelated healthy blood donors from the South Korean population were recruited through the same institution.

2.2.3. Sequencing samples

For the sequencing of genomic DNA, the DNA was extracted from peripheral blood lymphocytes as described in the DNA isolation section and arrayed into a 96 well plate format. The amount of DNA was 5 ng either liquid or dried to the well. The samples were unrelated German individuals affected with IBD from the population sample for association studies. A standard population sample for sequencing consisted of 30 unrelated German individuals affected with IBD (15 with CD and15 with UC) characterised through the criteria mentioned above. They were sampled thrice onto a 96 well plate each time with two empty control wells. A second IBD population sample for sequencing was somewhat larger with 47 unrelated German individuals (24 with CD, 23 with UC) and one empty control. They were applied to a 96 well plate twice. Initially the first cohort was used for sequencing of candidate genes. In situations of uncertainty caused by a bad sequencing result or only one individual showing a new polymorphism the second cohort was sequenced for confirmation of the results seen. Later the larger cohort became the standard sequencing sample. cDNA from 27 IBD cell lines from peripheral blood was obtained from the Department for Molecular Genetics, Max-Delbrück-Center for Molecular Medicine, Berlin. From this cDNA 5 ng were used liquid for sequencing.

2.2.4. Population samples for the analysis of LD structure

This study employs populations from 5 different ethnical backgrounds, all unrelated and not affected by IBD. The initial screening was carried out in a sample of 45 white US Americans who described themselves as being of European ancestry. These individuals were obtained from the Coriell Institute for Medical Research (Camden, NJ, USA) where they had been recruited as part of the Human Variation Panels. The subsequent in-depth analysis was performed in the 550 unrelated healthy German individuals employed as controls in the association study. Additional populations included (i) 93 unrelated Norwegian blood donors recruited at the Rikshospitalet Oslo, Norway, (ii) 78 unrelated white British individuals from the European Collection of Animal Cell Cultures, ECACC, Wiltshire, and (iii) 45 self-described US African-Americans also provided by the Coriell Institute.

2.3. Handling of samples

Patient recruitment and sample/data handling was in accordance with approved procedures by the local ethics committees at participating institutions. IBD ascertainment criteria were determined prior to the initiation of patient collection according to standard diagnostic criteria (57, 119). In cases of uncertainty of the diagnosis the patient was withdrawn from the study. Patient contact was initiated through the general practitioner or a medical centre and informed written consent was obtained from all participants. Besides two 9 ml EDTA blood vials a questionnaire inquiring disease circumstances, other diseases, and environmental factors was received from every patient and their family members (appendix 8.4). The blood was taken by the general practitioner and the questionnaire completed either by the participant or with the aid of the general practitioner. All material was sent back by the postal way (30, 93). The blood was immediately frozen to –80°C and stored at this temperature until the preparation of the DNA. Samples that were collected by collaborators were either sent as frozen EDTA blood as extracted DNA. Informed written consent was obtained as well from the non IBD sample populations.

2.3.1. DNA isolation

DNA isolation from EDTA-full blood by guanidine-detergent lysis with DNAzol® was performed either from fresh or from frozen samples, stored at -80°C (122). The original protocol was slightly adapted to optimise the yield of DNA from the samples used. About 9 ml blood was used. Frozen samples were thawed at room temperature immediately before preparation by gently inverting the blood vial occasionally. All following procedures were conducted on ice if not indicated otherwise. After the transfer to labelled sterile 50 ml tubes 18 ml (double volume) of MRC buffer (Molecular Research Center; 320 mM saccharose, 5 mM $MgCl_2$, 1 % Triton-X, 1 mM Tris at pH 7.5) was admixed and incubated for 10 minutes. During this step the red blood cells were destroyed and protein particles were separated from intact leukocytes. MRC buffer was made on a weekly basis according to the needs. The samples were centrifuged for 10 min at 10000 x*g* and 4°C. The supernatant was discarded and the pellet rinsed with 5 ml MRC buffer, which was discarded directly afterwards. After resuspension of the pellet in 9 ml MRC buffer the volume was refilled to 18 ml with MRC buffer. Incubation and centrifugation were repeated as described. This procedure was repeated until the pellet was white, containing only leukocytes. The pellet was resuspended in 5 ml of

DNAzol® and incubated at room temperature (r.t.) for 15 min (up to 30 min) until the solution was transparent. In this step the leukocytes were destroyed and the DNA was set free from the nuclei. Where this was problematic, proteinase K (stable for 4 weeks at -20°C) was added (100 µg/ml final concentration) with incubation at r.t. for up to 12 h for enzymatic digestion of proteins. 2.5 ml ice cold absolute ethanol (0.5 vol. per 1 vol. DNAzol®) was added and gently mixed by inverting the tube. A small thread of DNA precipitate floating in the liquid was transferred into a labelled DNAse free 1.5 ml tube. For low amounts and degraded DNA the sample was centrifuged for 5 to10 min at 5000 x*g* at r.t., the supernatant discarded and the pellet transferred. The precipitate was then washed with 96% ethanol *p.a.*, the liquid was discarded followed by a brief wash with 70% ethanol *p.a.*, discarding the liquid again immediately. The precipitate containing the DNA was then dried at r.t. for several hours (overnight) and then dissolved in 500 – 1000 µl Tris EDTA buffer (TE), depended on the approximate amount of DNA gained for 1 day at r.t. and afterwards at 4°C.

With a higher degree of throughput and automation of the DNA isolation process the protocol for the standard volume of 9 ml EDTA full blood was changed to the Invisorb Blood Giga Kit, which contains all chemicals used (Invitek, Berlin, Germany), following the protocol provided with the Kit. Frozen blood samples were thawed in cold water, the lid not touching the water, while gently inverting the blood vial occasionally. Thawed blood samples were transferred to labelled sterile 50 ml tubes previously filled with 30 ml of "Buffer 1" (4°C), shaken and incubated for 10 min at r.t.. During this step the red blood cells were destroyed while the leukocytes stayed intact. The samples were centrifuged for 3 min at 6000 x*g* and the supernatant discarded. Another 20 ml of "Buffer 1" one was added and the vial stirred until the pellet was dissolved followed by a second centrifugation as above. This step was repeated until the pellet was white, then the pellet was resuspended in 3 ml "Buffer 2" and 50 µl protein-kinase K followed by 2 h incubation at 60°C while rocking at 95 turns/min. The transparent solution with then free DNA (if not, the incubation was extended for another 1/2 h) was transferred to labelled 15 ml tubes, 1.8 ml "Buffer 3" were added and after stirring it was incubated for 5 min on ice, followed by centrifugation for 15 min at 10000 x*g*. The supernatant was transferred to a new, labelled 15 ml tube and the double volume (9.6 ml) of 96% ethanol *p.a.* added. Inverting the tube resulted in precipitation of the DNA, if precipitation did not take place the tube was incubated for 2 h at -20°C. After centrifugation for 3 min at 10000 x*g* the supernatant was discarded, the pellet transferred to a labelled 2 ml tube previously filled with 1 ml 70% ethanol *p.a.*, stirred and centrifuged for 2 min at 10000

x*g*. The supernatant was discarded, the pellet dried in the tube with open lid until the ethanol was completely evaporated. Finally 500 µl TE (1x) buffer was added to dissolve the DNA.

For small volumes of blood DNA was extracted with the PEQLAB DNA isolation system (PEQLAB Biotechnology GmbH, Erlangen, Germany). All procedures were conducted according to the protocol provided with the chemicals and at room temperature if not indicated otherwise. About 1.5 ml EDTA full blood was transferred to a labelled 15 ml reaction tube. 150 µl OB^{TM} protease and 1.5 ml "BL buffer" (all buffers and chemicals except for isopropanol and ethanol *p.a*. were part of the PEQLAB DNA isolation system) were added and stirred. During 10 min incubation at 70°C, the samples were stirred once. 1.56 ml isopropanol was added and again stirred. At this point all cell components were destroyed and the DNA was released. The complete liquid was divided to two HiBind_DNA columns placed in two 2 ml collection vials and centrifuged each time 750 µl for 1 min at 8000 x*g*. While the DNA binds to the silica-filter of the column, other cell particles were washed through. Flow through and collection vial were discarded and the columns transferred to new collection tubes. 750 µl wash buffer (completed with 1 1/2 the volume 100% ethanol *p.a*.) was added to each column, and then centrifuged 750 µl for 1 min at 8000 x*g*. The wash step was repeated, the flow through discarded but the collection tube was used again for 2 min centrifugation at 8000 x*g*. During this procedure protein residues were removed from the filter of the column. Both columns were placed into freshly labelled 1.5 ml reaction tubes. 200 µl elution buffer (70°C) was added on each column matrix followed by incubation for 2 min at r.t. to elute the DNA from the silica-filter. The columns were centrifuged at 8000 x*g* for 1 min, the flow through recovered and transferred again to the column. After another incubation of 2 min the samples are centrifuged again at 8000 x*g*. for 2 min. The DNA then was ready to use.

DNA samples were stored either at 4°C or at -20°C. The quality and the concentration of the DNA were measured when all DNA was dissolved. If there was still precipitate in the tube it was centrifuged for 1 min at 8000 x*g* (r.t.) and the supernatant transferred to a new, labelled tube and measured, while the original tube was refilled with 500 µl TE (1x) buffer. The genomic DNA was quantified using PicoGreen (Molecular Probes Europe BV, Leiden, The Netherlands). The automated concentration measurement was carried out on a TECAN SPECTRO FLUOR fluorescence microplate reader (Tecan Deutschland GmbH, Crailsheim, Germany).

2.3.3. Application to 96 and 384 well format

Individual DNA samples were arrayed in 96 well microtiter plates. Each 96 well plate contained a maximum of 93 DNA samples. Two wells contained the same CEPH (Centre d'Etude du Polymorphisme Humain, Paris, France) cell line (as a control in the diallelic discrimination assays) positioned for each plate layout at the same position, and initially 1, later 4 wells with always the same position were no DNA containing empty controls. Four 96 well plates were arranged to one 384 well plate. Wherever family samples were involved the plates were designed to keep the families on one plate. The plate-layout with individual identification through barcodes was entered to the database system before the plate was produced. Each plate received an identification number. Application to 96 and 384 deepwell plates (ABgene, Epsom, UK) and adjustment of the concentration was performed with the aid of a TECAN Genesis RSP 150 multipipetting robot (Tecan Deutschland GmbH, Crailsheim, Germany). An aliquot of DNA -TE solution containing 2 ng (for diallelic discrimination assays) or 5 ng (for sequencing) DNA was then distributed via 96- and 384-channel Robbins Scientific Hydra microdispensers (Dunn Labortechnik GmbH; Asbach, Germany) to the 96 or 384 well microplates (Costar Corning Incorporated; Cambridge MA, USA, Greiner Bio-One GmbH, Frickenhausen; Germany) to be dried at 60°C, the plates then were heat sealed with a ABGENE ALPS 300 (ABgene, Epsom, UK) for storage.

2.4. Diallelic genotyping

2.4.1. The principle of diallelic genotyping

For the analysis of allele status at a known SNP position in a large group of DNA samples the method of diallelic genotyping (5’ nuclease assays) was applied. The 5' nuclease PCR assay detects the accumulation of specific PCR product by hybridisation and cleavage of a double-labelled fluorogenic oligonucleotide during the amplification reaction. The method is based on a PCR reaction including the SNP of interest. Besides the forward and reverse primer, 2 oligonucleotides (called probes), one for each allele, are involved in the reaction. The probe has a reporter fluorescent dye attached at the 5' end and a quencher dye at the 3' end. The two probes have reporter dyes with fluorescent emission at different wave-length. During the PCR reaction the oligonucleotide with the correct nucleotide at the position of the SNP hybridises to the reverse DNA strand and is cleaved through the 5’ nuclease activity of Taq DNA

polymerase. An increase in reporter fluorescence intensity indicates which probe has hybridised to the target PCR product and has been cleaved (123). Through separation from the quencher the dye becomes fluorescent (Fig. 2.1). Visualisation takes place through laser scanning technology (ABI Prism® 7700 Sequence Detection System or ABI Prism 7900HT Sequence Detection System). Fluorescence increases in each cycle, proportional to the rate of probe cleavage. To induce fluorescence, laser light is distributed to the 96 or 384 sample wells via a multiplexed array of optical fibres. The resulting fluorescent emission returns via the fibres and is directed to a spectrograph with a charge-coupled device (CCD) camera. The fluorescent dyes employed for the analysis are TET (tetrachloro-6-carboxyfluorescein) and FAM™ (6-carboxyfluorescein) and VIC® (trade name by Applied Biosystems) and the quencher TAMRA (6-carboxytetramethylrhodamine). The systems employed (Applied Biosystems TaqMan® technology und TaqMan®-MGB technology) allow 96 well and 384 well platforms. The integrated software depicts the fluorescent status for each well in a diagram and allows to assign an allele calling to each well. A general setting of the software is that the TET (VIC® for TaqMan®-MGB probes) allele is always called 1, FAM™-allele 2. Each well of the 96 well or 384 well plate results in a dot on the diagram according to the fluorescent intensity of the reporter dyes. The distribution is a cloud of the dots for each homozygous allele 1/1 or 2/2 calling group, and the heterozygous 1/2 calling group. One homozygous cloud is usually large while the heterozygous cloud is of medium size and the other homozygous cloud is small. For SNPs with one allele at a low frequency the smallest cloud could be missing sometimes. In cases where the SNP is not polymorph in the individuals of the plate analysed, only one cloud is present in the diagram.

In contrast to the TaqMan® probes the TaqMan®-MGB probes contain a minor groove binder (MGB) which attaches to the minor groove of duplex DNA and by this stabilises hybridisation products and allows shorter probes. Furthermore a non-fluorescent quencher replaced the TAMRA™ quencher, reducing the background fluorescence and improving the spectral discrimination.

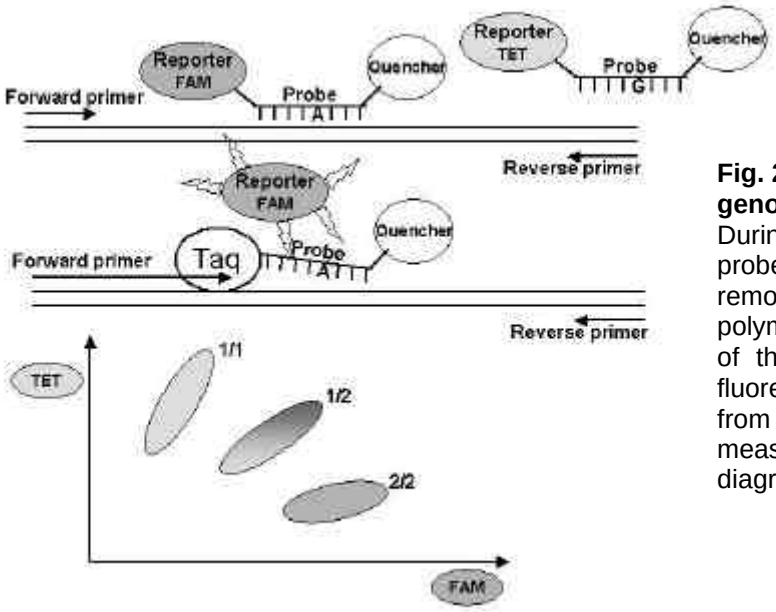

Fig. 2.1: The principle of diallelic genotyping.
During the PCR the compatible probe anneals to the DNA and is removed through the Taq-polymerase during the elongation of the primer. This activates the fluorescent dye and the emission from each sample can be measured and plotted in a diagram.

2.4.2. The method of diallelic genotyping

Different levels of automation were employed in the diallelic genotyping. Many allelic genotyping assays were ordered in a ready to use status (Assay-on-Demand, Applied Biosystems Inc., Foster City, CA, USA), or were designed from a sequence by a company (Assay-by-Design Applied Biosystems Inc., Foster City, CA, USA) (both TaqMan®-MGB technology). Both need no further optimisation. Several of the assays employed were self-designed using the TaqMan® technology. The oligonucleotides were manufactured and labelled with fluorescent dyes by a company (Eurogentec S.A., Seraing, Belgium) according to the defined sequence.

For the design of oligonucleotides employed in the diallelic assay the program Primer Express 1.0 or 2.0 (Applied Biosystems Inc., Foster City, CA, USA) was used. The guidelines proposed in the program description were applied (124-126). The primer concentration was set to 50 nM. In the TaqMan probe document the probe was selected manually. The polymorphism was in the middle or the last third of the probe (minimum 5 bases before the end of the probe). The probe was not allowed to start with a guanine (G) as the first base. The required melting temperature (Tm) for the probes was about 70°C, a range of 2°C up and down is possible). The difference in Tm between the probes was not larger than 0.5°C. The sequence of the probes belonging to one assay was allowed to vary in length, with a

maximum length of 40 bases, the minimum was 17 bases. The primer Tm was about 10°C below the Tm of the probes, therefore the optimal Tm for primers was between 58 to 60°C. Primers were not selected from repeat masked regions or regions with other SNPs. The average amplicon length varied between 50 to 150 bp, where the SNP needed verification the amplicon was chosen close to 150 bp or even longer to allow a clean sequencing product around the SNP. The program proposed up to 200 primer pairs for the selected probe. The forward primer was chosen at least at a 10-20 bases distance from the probe. For primers used for sequence validation minimum 40 bases distance to the SNP was required, eventually the reverse primer was used for sequencing. The reverse primer was allowed to overlap with the probe, but not to reach the SNP. Primers with high variation in the bases at the 3'end, no T at the last position, no three same bases in a row in the last third of the oligonucleotide were allowed. If the program was not able to find primers, the Tm and the amplicon length were altered slightly or a new probe was selected. In the primer test document the primers were tested for self-annealing and loops. Not more than three self-annealing or loop-bonds in a row were accepted, primer-dimerisation for up to 4 bonds was allowed. TaqMan® probes were labelled with the fluorescent dyes FAM™ or TET and with the quencher TAMRA™.

For TaqMan® probes an optimisation of the primer concentration was needed. The optimisation was adapted from the procedure proposed in the protocol for the TaqMan® allelic discrimination (Applied Biosystems Inc., Foster City, CA, USA)(125) to a high throughput level including several control mechanisms. Therefore primers and probes were diluted to 100 µM storage concentration with double distilled water (DDW), work dilutions were at 20 µM for the primer and 10 µM for the probes. By adjusting initial concentration the effective Tm of the primer shifted by –2°C up to +2°C from the midpoint. This test was usually performed with only the FAM™ labelled probe. The following reaction concentrations were tested for each primer: 50 nM, 300 nM 900 nM, the concentration 50/50 was not tested in the optimisation (the reaction solution for each concentration is shown in Table 2.4). Each concentration combination was tested 3 times with 2 ng DNA (dried) and on 3 blank controls. The final concentration for the FAM™ probe was 100 nM. The reaction volume for the optimisation was 25 µl, with 16.6 µl of the TaqMan® Universal PCR MasterMix (Applied Biosystems Inc., Foster City, CA, USA). The standard cycle conditions was 60°C with 45 cycles on a GeneAmp® PCR System 9700 (Applied Biosystems Inc., Foster City, CA, USA).

Table 2.4: Reaction solution for the probe concentration optimisation.
DDW = double distilled water

forward/ reverse primer (nM final conc.)	Mix (µL)	FAM™ probe (10 µM) µl	forward primer (20 µM) µl	reverse primer (20 µM) µl	DDW	total volume µl
50/300	16.6	0.33	0.083	0.5	15.75	33.3
50/900	16.6	0.33	0.083	1.5	14.75	33.3
300/50	16.6	0.33	0.5	0.083	15.75	33.3
300/300	16.6	0.33	0.5	0.5	15.33	33.3
300/900	16.6	0.33	0.5	1.5	14.33	33.3
900/50	16.6	0.33	1.5	0.083	14.75	33.3
900/300	16.6	0.33	1.5	0.5	14.33	33.3
900/900	16.6	0.33	1.5	1.5	13.33	33.3

After the PCR reaction the fluorescence was detected with the ABI 7700 Sequence Detector (Applied Biosystems Inc., Foster City, CA, USA) and the ΔRn for DNA and blank controls was measured. While Rn is the emission of the normalised reporter dye, Rn+ is the fluorescent emission when a DNA template is present and Rn- is the emission measured without a template, the ΔRn is calculated from (Rn+) - (Rn-). The primer concentration with the greatest difference between no template control and the DNA template and a constant value over all three samples was chosen. The assay was then tested on a plate with different DNA samples at the annealing temperatures 58°C, 60°C and 62°C with the final reaction volume of 5 µl (for earlier assays the reaction volume was 10 µl). Where no sufficient differences were observed the reaction was repeated with additional 2.5 µl $MgCl_2$ (25 µM). The standard protocol for the assays is shown in the Tables 2.5 and 2.6.

For SNPs that needed sequence verification the assay was tested on the sequenced samples. Other quality criteria were clearly separated clouds of dots in the analysis program, a small number of not amplified samples, no cloud duplications (more than 3 clouds total), no well with empty control in one of three calling dot clouds (otherwise contamination had happened), and all CEPH control DNA samples needed to be in the same cloud of dots. Hardy-Weinberg Equilibrium (HWE) was calculated for the independent DNA samples. After all samples for one assay were genotyped and analysed, the data were imported to the internal database system where additional quality tests were performed.

For a higher automation on the 384 well platform a TECAN Genesis RSP 150 multipipetting robot (Tecan Deutschland GmbH, Crailsheim, Germany) was used to apply the complete reaction solution to the 384 well plates with the dried DNA. The PCR reaction was done on

T1 Thermocycler with 384 well application (Whatman Biometra GmbH, Göttingen, Germany) or GeneAmp® PCR System 9700 with 384 well application (Applied Biosystems Inc., Foster City, CA, USA). Plate fluorescence was measured automatically with an ABI Prism 7900HT sequence detection system (Applied Biosystems Inc., Foster City, CA, USA). Allele calling for each plate was done manually to ensure data quality. All markers had to have a genotyping performance of over 95%. Markers with lower performance were not included into the analysis.

Table 2.5: Reaction solution for the diallelic genotyping PCR.

For a 5 µl PCR reaction volume	Volume (µl) for one reaction	Final concentration (nM)
TaqMan® Universal PCR MasterMix	2.5	---
FAM™ probe (10µM)	0.05	100
TET probe (10µM)	0.05	100
Forward primer (20µM)	0.0125/0.075/0.225	50/300/900
Reverse primer (20µM)	0.0125/0.075/0.225	50/300/900
Water	to 5 µl total volume	---

Table 2.6: Temperature cycling conditions for the diallelic genotyping.

Temperature	Time	process	function
50°C	2 min	hold	optimal AmpErase UNG activity
95°C	10 min	hold	DNA polymerase activation
95°C	15 sec	cycle 40X	melting
58/60/62°C	1 min		annealing and extension
10°C	∞	hold	

2.4.3. SNP marker selection

For the association mapping study of the 20 Mb large peri-MHC (microsatellite markers D6S461 to D6S271) region SNPs were selected from the NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP/index.html) or from literature. The SNP markers were selected to achieve an average density of one marker every 50 kb. Selection criteria for SNPs from dbSNP included that the SNP was tested in 10 or more chromosomes and the description was in genomic sequence, the minimum required sequence length was 120 bp on each side of the SNP. Polymorphisms were preferentially selected from coding sequence causing a change in the amino acid sequence. To achieve a tolerable density of the SNP marker, polymorphisms were chosen as well from intronic, untranslated or even anonymous DNA sequence.

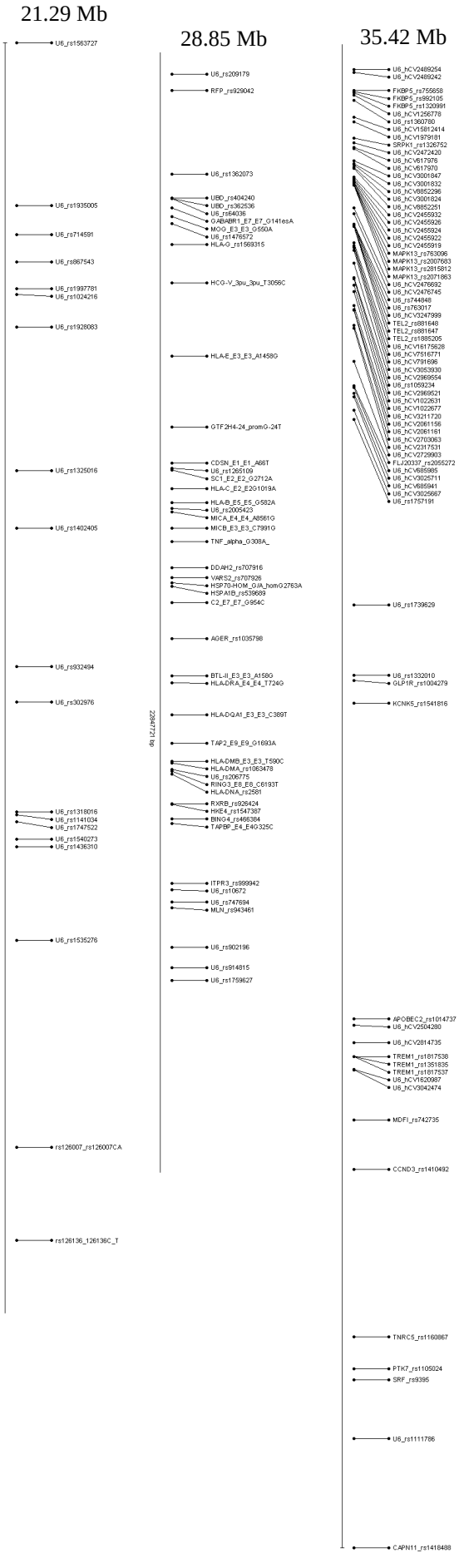

Fig. 2.2: SNP marker map for association mapping on chromosome 6p21.
SNP marker density was higher in the central part covering the MHC region than in the telomeric and centromeric end of the map.

SNPs described by "The Sequencing Consortium" (http://snp.cshl.org/) were preferred as they were always described in genomic sequence, other SNPs in the direct environment were indicated and repeat masked sequence was identified. All SNPs from the dbSNP were sequence verified prior to genotyping, tested on the sequenced samples for correct calling and had to have a genotyping performance of over 95%. Further quality standards were no significant derivation from the HWE in the controls and a test for Mendelian inheritance errors. A total of 142 SNP markers (self designed, Assay-on Demand and Assay-by-Design, Applied Biosystems, Foster City, CA, USA) were included in the analysis after quality control. Physical positions of the SNPs were determined from the NCBI assembly (build 32). The density of the markers was not equally distributed. The average density was 161 kb with a median of 60 kb. The maximum distance between two markers was 1860 kb, the shortest distance 68 bp. While a central area from 29.4 Mb to 37.5 Mb including the MHC region had a higher density (average of 80 kb, median 36 kb), the outer areas from 21.3 Mb to 44.1 Mb were less densely covered with markers (Fig. 2.2). Genotyping was performed in the 550 control individuals, in the 178 multiplex families and in the 461 monoplex families.

In order to explore the structure of linkage disequilibrium existing in the region around the MHC diallelic genotyping (5' nuclease) assays were selected for SNPs located in a 10 Mb region on human chromosome 6p (microsatellite markers D6S461 to D6S291) between position 25.3 Mb - 36.3 Mb (NCBI release 30). SNPs were chosen from the Celera Human RefSNP database (version 3.6) through a prioritisation scheme that stipulated evidence for an independent discovery of the minor allele (127). PCR primers and TaqMan®-MGB probes were designed using an algorithm that generates oligonucleotide sequences and subsequently screens them against a genome sequence database, to avoid potential genotyping artefacts. SNPs were selected by imposing a "picket-fence" with an average grid density of 10 kb upon the genomic sequence of interest (127). Only those 920 SNPs that were found to be polymorphic in 10 unrelated DNA samples were included in subsequent experiments (Assay-on-Demand^{IM}, Applied Biosystems, Foster City, CA, USA). The full 920 SNP markers were typed in the 45 white US American individuals from the Coriell sample.

Consequently to the analysis of the large region a smaller region representing all characteristics of LD observed in the initial screening was selected from the initial area. Being as well of special interest in the context of disease gene detection, a 3.53 Mb region of chromosome 6p21.3 covering the MHC (NCBI release 30 position 29.4 Mb - 32.9 Mb)

representative in this sense, was chosen for a more detailed and comparative analysis of haplotype structure in five different populations. An additional 37 SNPs were obtained from the association study marker set and another 11 SNP markers from external sequencing (Institute of Immunology University Clinic Schleswig-Holstein, Kiel, Germany) were included in order to enhance marker spacing and distribution of the 272 SNPs located in this region from the first set. Physical positions of the total 320 SNPs were determined from the NCBI assembly (build 30). The final average marker density in the region was 12 kb (median 5.5 kb), with a minimum of 14 bp and a maximum of 189 kb. These SNP markers were typed in the 550 German, 93 Norwegian, 73 UK and 45 US African-American individuals (Coriell). Genotyping of the Coriell samples was conducted at the Services Development and Delivery Laboratories of Applied Biosystems, Foster City, California, USA.

2.4.4. Genotyping of HLA-DPA1

The HLA-DPA1 was one of several possible candidate genes in the HLA class II region. It is expressed on intestinal epithelial cells of non-inflamed small and to a lesser extent large bowel (128). Expression is increased during intestinal inflammation in CD and UC (128, 129). Increased expression on epithelia in UC appears associated with the number of infiltrating mononuclear cells (130). In moderate to high inflammatory activity HLA-DP antigens were expressed strongly in enterocytes, glial cells, and capillary and venular endothelium (131). Increased expression of HLA-DP could be induced *ex vivo* by interferon-gamma (IFN-γ) and tumour-necrosis factor alpha (TNF-α) in isolated enterocytes and colonic epithelial cells from jejunal mucosal biopsies (132-134).

Instead of the assessment of all HLA-DPA1 allelic variants known, polymorphisms at the nucleotide positions 91-92, 111, 114 and 149 (of the coding sequence) in exon 2 were analysed through sequencing. Exon 2 is relevant in the formation of the binding cleft and an alteration could influence the antigen presenting function. Primers for PCR amplification of the second exon were designed on sequences obtained from the NCBI Database (Accession Number X03100, Human HLA-SB (DP) alpha gene). The forward primer (5'TCAGGATGCCCAGACTTTCAA) was chosen from intron 1, the reverse primer (5'CAGGGGGCACTTAGGCTTCC) from intron 3. The oligonucleotides used as internal primers for sequencing (forward: 5'GCGGACCATGTGTCAACTTAT, reverse: 5'GCCTGAGTGTGGTTGGAAG) were adopted from the literature (135) while the

optimised PCR and sequencing conditions were developed on the basis of PCR standard protocols (124, 136). Amplification was performed in 35 cycles at 60°C annealing for 30 sec. and 1.30 min elongation at 72°C. (AmpliTaq Gold Standard chemistry) with a GeneAmp® PCR System 9700 (both Applied Biosystems Inc., Foster City, CA, USA). Enzymatic purification of PCR products and sequencing were performed according to the standard sequencing protocol (see section 2.5.3). Obtained sequences were compared with Sequencher (Gene Codes Corporation, Ann Arbor, MI, USA) to consensus sequence (accession number X03100.1) and allelic variants. The three polymorphic sites were handled like the SNP markers with 2 alleles for the position 91-92, three alleles at position 111-114 and two alleles at position 149 (see as well Table 3.6). Allele variant data were entered to the database and subjected to quality control and analysis alike the diallelic genotyping results. Analysis was performed in the German family and trio IBD sample population, the South African family sample population and independent IBD cases and control samples from South Korea.

2.5. Mutation detection and verification of SNP markers

2.5.1. PCR optimisation

The strategy for the optimisation of PCR conditions for different primer-pairs was developed on the basis of standard PCR protocols (136, 137). Adaptations were made in order to render a high throughput possible and to maintain quality standards. After dilution of the oligonucleotide to a storage concentration of 100 pmol/µl with DDW, a working solution of 20 pmol/µl was made. The reaction solution contained the following reagents: 2.5 µl Qiagen Buffer (10 x concentrated) 0.5 µl dNTP (10 mM), 0.2 µl primer forward; 20 pmol/µl, 0.2 µl primer reverse; 20 pmol/µl 1.0 µl $MgCl_2$ (25 mM) 0.15 µl Taq polymerase (QIAGEN GmbH; Hilden, Germany) and DDW, to a final amount of 25.0 µl per reaction. Two ng DNA were used, either already dried in the plate or if liquid DNA was added the amount of water was reduced by the equivalent volume. Optionally there was the possibility to add Qiagen special buffer for problematic amplifications. Also the amount of $MgCl_2$ could be changed according to the conditions needed. The reaction solution was set up for 12 reactions (of which 10 were used). The PCR reaction was set up in a T1 Gradient (Whatman Biometra GmbH, Göttingen, Germany) with the following cycle program: 96°C - 10 min, (96°C - 1 min; 64°C - 1 min {– 0.5 °C per step}; 72°C - 1 min) 16x, (96°C - 1 min; 56°C - 1 min; 72°C - 1 min) 15x, 72°C - 10 min, 10°C ∞. The gradient covered the temperatures up to 5°C higher and lower than the

annealing temperature given in the protocol. This way a range of temperatures can be tested at once (e.g. a range from 59°C to 69°C in the example above). If needed, the temperature was altered to find the optimum for a primer combination. Using a 96 well plate up to 8 primers were tested at once. Five µl PCR products together with 1 µl 6-times loading buffer (0.25% bromophenol blue, 0.25% Xylene Cyanol FF, 30% Glycerol in Water) were applied to a 1.5% agarose gel (300 ml Tris borate EDTA (TBE), 3 µl ethidium-bromide). A 100 bp ladder (Invitrogen GmbH, Karlsruhe, Germany) was applied next to the PCR products, electrophoresis conditions were 150 V for 50 min. A picture was taken under UV light and observed bands of PCR products were compared to expected product length. Quality criteria were one clear band at the correct length without smear, no double bands, and a low amount of primer-dimer. If the quality criteria were not met several factors were open for alteration ($MgCl_2$ concentration, adding Qiagen special buffer, temperature change and alteration of primer concentration).

2.5.2. Sequence analysis

After primer optimisation a reaction solution was prepared according to the conditions optimised before. The solution was applied to a 96 well plate (depending on the sequencing population occupying 48 wells or 32 wells) with 5ng DNA either dried to the wells or liquid. The plate design allowed PCR amplification for 2 or 3 primer-pairs at the same time if they had the same amplification cycle conditions. Five µl PCR product were mixed with 1 µl 6-times loading buffer and tested on a 1.5% agarose gel (in 300 ml TBE, 1% ethidiumbromide, 150 V for 50 min). Unwanted primer-dimers, and free dNTPs in the PCR product were removed by digestion. For a highly concentrated PCR product a 1:5 with DDW dilution was necessary before the digest, which was decided from the appearance of the bands on the agarose gel under UV light. If the bands were bright, the PCR product was diluted, if they were weak the original concentration was used. The enzymatic digest was performed in a new plate with 8 µl PCR product (diluted) at the following conditions: 0.30 µl SAP (Shrimp Alkaline Phosphatase; 1 U/µl) 0.15 µl ExoI (Exonuclease I; 10 U/µl) 1.55 µl DDW adding up to 2.0 µl complete reaction mix, with following incubation conditions: 37°C - 15 min, 72°C - 15 min, 4°C - ∞, adapted from "the Current Protocols in Human Genetics" **(138)**. The digested PCR product was again analysed through electrophoresis to test for the cleanness of the sequencing template. The chemicals used for sequencing were from the BigDye Terminator Ready Reaction kit (Applied Biosystems Inc., Foster City, CA, USA) based on

fluorescent truncation dNTPs. Two µl of digested PCR product were used for the sequencing reaction, 1.0 µl primer (forward); 1.6 pmol/µl, 1.0 µl BigDye Ready Reaction Mix from the kit, and 6.0 µl DDW for a reaction volume of 8.0 µl. The same reaction was performed with the reverse primer. The primers used in the sequencing reaction were either the same as from the PCR reaction or nested primer. The following cycle protocol for the sequencing reaction was used: 95°C - 5 min; (95°C - 1 min; 50°C - 45 sec; 60°C - 4min) 25x; 60°C -.5 min; 10°C - ∞. Sample cleanup was performed with Sephadex spin columns. The Sephadex spin column plates (MAHVN 4550) were prepared by adding Sephadex powder (G50 Superfine) with the aid of a Multiscreen Column loader (all through Amersham Biosciences, Freiburg, Germany), 300 µl DDW were added to each column. After incubation for 3 h at room temperature with closed lid, the spin column plate was centrifuged at 910 xg for 5 min to eliminate the overdue water. Again 150 µl DDW was added followed by immediate centrifugation at 910 xg for 5 min. The sequencing product was replenished with DDW to 30 µl total volume and pipetted to the centre of the spin column, the plate was fitted to a MicroAmp® Optical 96-Well Reaction Plate (Applied Biosystems Inc., Foster City, CA, USA). After centrifugation at 910 xg for 5 min, the flow through contained the purified sequencing product. The sequence detection was performed with an automated, high-throughput, 96-capillary fluorescence detection system ABI PRISM® 3700 DNA Analyzer (Applied Biosystems Inc., Foster City, CA, USA), for fluorescent labelled DNA fragments. For the sequence analysis chromatograms were aligned and compared to the consensus sequences using Sequencher Version 4.0.5 (Gene Codes Corporation, Ann Arbor, MI, USA).

2.5.3. SNP verification

SNPs from the NCBI Database dbSNP needed verification through sequencing in a set of patients. In this case only primers were designed and ordered. Temperature and amplification condition optimisation were carried out as described above. Sequence verification of SNPs was usually performed in 23 IBD samples (plus one no DNA control) and only on one strand. This allowed the verification of 4 SNPs at once where the PCR cycle conditions were compatible. The 96 well plate designed for this purpose contained 4 times the same set of samples. Otherwise the sequencing protocol was followed as described above. The sequencing data were compared to the sequence in the SNP description in dbSNP or, where available, the TSC database (http://snp.cshl.org/) using Sequencher Version 4.0.5 (Gene Codes Corporation, Ann Arbor, MI, USA). The allelic setting for each sample was noted.

Additional SNPs in the direct environment were identified as the design of probes was affected through them. For SNPs that are not polymorphic in this set of patients the SNP was either eliminated from the list or sequenced in a second set (Sequencing population 47 samples). SNPs that were not polymorphic were excluded. After optimisation, the SNP-assay was tested on the same set of samples. Allele-calling needed to be identical to the sequencing results, if not the assay needed further optimisation until the allele-calling was identical, or the SNP was excluded from further analysis.

2.5.4. Mutation detection in candidate genes

Several genes qualified themselves as candidates for association with IBD through their function and their position within the linkage region and near to association peaks. These genes were sequenced for polymorphisms in the exons. For all genes an alignment of mRNA or cDNA sequence with genomic sequence was performed to determine the exact positioning of the exons and to identify the known splice variants. Primers were designed on intronic sequence and the analysis performed in genomic DNA. Where nested primers were employed for sequencing, they were designed at minimum 40 bases distant from the exon boundaries for all exons. Mutation detection was performed in 30 unrelated German individuals affected with IBD (15 CD and 15 UC) or 47 unrelated German individuals affected with IBD (24 patients with CD, 23 patients with UC). PCR amplification and sequencing were done according to the protocol described. Experimental sequences were aligned with consensus sequence using Sequencher Version 4.0.5 (Gene Codes Corporation, Ann Arbor, MI, USA). In case of uncertainty sequencing was repeated in an additional group of individuals. For newly found exon polymorphisms diallelic genotyping assays were designed and tested in the full German association study sample population. Positioning of the genes is given in the NCBI assembly build 32. Primer oligonucleotides used for PCR amplification and sequencing reaction are listed in the appendix 8.2, Table 8.3.

2.5.4.1 MAPK14

Three splice variants are well described for the MAPK14 gene locus, identified by the names MAPK14 (or p38α), Mxi2 and CSBP1. An alignment of cDNA sequences of MAPK14/p38-α (accession number L35253), Mxi2 (accession number U19775) and CSBP1 (accession number L35263) with the PAC genomic clone 179N16 (accession number Z95152), generated from chromosome 6p21.1/21.33, revealed 13 exons and the complete intronic

region of the p38-α gene as described previously (139, 140). The complete coding sequence is positioned from 35990 kb to 36075 kb on chromosome 6p21. Exons herein are numbered according to this alignment resulting in an alternative splicing of exon 9 and 10 for MAPK 14 and CSBP1 and an extended exon 11 for Mxi2 (Fig. 3.6). While the amino acid sequence is identical for position 1 to 280 (exons 1 to 8), alternative internal splicing of exon 9 and 10 accounts for 14 amino acid exchanges between MAPK14 and CSBP1. Mxi2 lacks 80 COOH terminal amino acids (exons 12 and 13), which are replaced by 17 residues encoded by exon 11' (contiguous to exon 11) (140, 141). The variants MAPK14 and CSBP1 are described to be expressed at different levels in most human tissues (142-144) while in mouse Mxi2 is expressed solely in kidney (140). The relevance in inflammatory processes and the position within the chromosome linkage region suggested the p38-α gene locus an interesting target for mutation detection and further analysis in IBD affected individuals. The protein encoded by this gene is a member of the MAP kinase family. MAP kinases act as an integration point for multiple biochemical signals in the cell signalling process. Various environmental stresses and proinflammatory cytokines cause activation to MAPK14. The activation requires phosphorylation by MAP kinase kinases (MKKs). It has an important part in stress related transcription and cell cycle regulation. All 13 exons were PCR amplified by distinct primer-pairs, furthermore the potential promotor region covering about 1000 nucleotides upstream of the 5'end of exon 1 was sequenced. All primers were tested negatively for homologous binding in p38-β, -γ and -δ. Sequencing was originally done in the set of 30 individuals. Where the results were not clear the sequencing reaction was repeated in the additional set of 47 IBD affected individuals.

2.5.4.2 MAPK13

For MAPK13 (or p38-δ) no additional splice variants were described in the literature. Twelve exons were confirmed in alignment of the mRNA (accession number AF004709) with PAC genomic clone 179N16 (accession number Z95152). The gene is positioned at 36092 kb to 36105.5 kb on chromosome 6p21 centromeric to MAPK14. It is closely related to MAPK14 and involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development. Activation is triggered through proinflammatory cytokines and cellular stress. All primers were tested negatively for homologous binding in p38-α, -β, and -γ. Sequencing was done in the cohort of 47 individuals (Fig. 3.6).

2.5.4.3 TREM1

The TREM1 gene locus encodes for a triggering receptor expressed on myeloid cells. The expression and functional properties of TREM1 suggests a role in acute inflammation (145). The gene is located at 41232 kb to 41251 kb on the chromosome 6p21. The mRNA (accession number AF196329) encodes for 4 exons when aligned with genomic DNA (accession number NT_007592). Primers for the exons and a potential promotor region (about 300 bp proximal of exon 1) were designed on genomic DNA and sequencing was performed in a cohort of 30 individuals (Fig. 3.6).

2.5.4.4 BRPF3

The BRFP3 gene was only predicted at the time it was analysed. The predicted transcript Enst00000259697 (ensemble: http://www.ensembl.org/) was aligned with chromosome 6 contig NT_007592.9, resulting in 12 pieces of mRNA matching the genomic DNA. The transcript has by now been confirmed. The gene function is still unknown, the encoded protein contains a bromodomain and PHD finger. It is positioned neighbouring centromeric to MAPK13. To confirm the existence of the transcript the amplification and sequencing was performed in cDNA from 27 cell lines derived from peripheral blood lymphocytes of IBD cases.

2.6. Internal database

2.6.1. Follow up of phenotype-genotype sample data

The large numbers of individuals with DNA samples, phenotype description, family information and genotype information was handled through an internal database system (146), which is an SQL Server database. All samples received an identification number at their entrance into the laboratory. Information about the individual like phenotypic trait (affected or unaffected, specification to UC and CD), gender, age, family relationship (pedigree information) and population information were entered to the database and linked to each other. Additionally, sample information like quantity and concentration of DNA and plate information (which patient's DNA is in each well of a 96 or 384 well plate) was connected. The plates with the DNA samples were again identified through a barcode system. Information about diallelic genotyping assays like oligonucleotide sequences (not available for Assay-on-Demand assays), chromosomal position, and genetic variants was entered as

well in the database. When an assay was applied to a sample and the resulting genotypes were entered in the database, the information was automatically linked to the individual sample through the barcode of the plate. The database allows to export data of individuals, within an analysis population, with their pedigree information, phenotypic trait and genotype information from a set of self selected SNP markers. The export format is the appropriate format for a selection of analysis programs (LINKAGE PREFILE). Besides that, a variety of application tools allowed to see which assays were genotyped on which samples (or plates), construct individual maps of markers together with the chromosomal position, and several quality measures. Additional applications helped to enter and edit sample information, create plate templates, to control robots to make plates and to measure DNA concentration.

2.6.2. Quality control

Integrated to the database system was a test for Mendelian inheritance errors (147). This application was part of the SNP genotyping data import to the database. It was utilized for each SNP marker and all populations the marker has been genotyped for. All inheritance errors were shown in a list with all family members and their genotyping data for that assay. The positions of the family members in the fluorescence diagram of the plate containing the family were seen through the linked plate-view (146). If the positions of one or more family members were not clearly within one of the clouds defining the genotype, the allele calling was wrong. The genotype data of the corresponding family members were set to 0. In cases where the positioning was clear, one of the individuals assigned to the pedigree was not a real member of the family. This might have been caused through a mix up of the samples or through infidelity. A few families with Mendel errors were normal and acceptable. The genotyping data for these were set to "0" and therefore were not included into the analysis. For SNP markers with a high number of inheritance errors, the SNP marker was excluded from the analysis when the assay was evaluated as not reliable. For sample populations of single individuals genotyping results were subjected to assessment of HWE by a χ^2-test at the 1% significance level. This test was applied as well to genotypes from families, however a slight unbalance could be expected in this situation. A quick view tool of the database showing the allele calling allowed to test if plates had been called inversely by accident. In such a situation the import had to be repeated with the correct calling.

2.7. Statistical analysis

2.7.1. Association analysis

Genetic analyses were performed using the diagnostic disease categories IBD, CD and UC as described. Allele and genotype frequencies for each SNP were calculated from all unrelated control individuals. For the calculation of the allele and genotype frequency in the cases only one affected individual per multiplex family was selected randomly, in order to avoid a bias through overrepresentation of a family genome. From the monoplex families each affected individual was included. This was done separately for each category IBD, CD and UC. On the basis the allele and genotype numbers received in this process case control statistics were performed for alleles and genotypes using contingency tables (43). Pearson's χ^2 was used as a measure of independence or association between the disease trait and the allelic variants of the SNPs tested or their genotypic prevalence in the case and the control samples. The significance at the 95% confidence interval is given as p-value, additionally a permutation test with 1000 repeats was performed to verify significance. Family based transmission distortion was analysed in the study population using the transmission disequilibrium test (TDT) from TRANSMIT version 2.5 (148). This test is based on a comparison of alleles transmitted from parents to affected offspring against the alleles that were expected to be transmitted from the population frequency in the parent generation according to HWE calculations. For each haplotype or allele, a test for excess transmission of that haplotype was performed. A bootstrap significance testing was carried out using 1000 bootstrap samples. The family based association was calculated as well using Genehunter (40). This program calculated the TDT according to the original description (42) by comparing alleles transmitted from parent to offspring with alleles untransmitted. Transmissions from homozygous parents were not informative in this analysis. In addition, the transmissions and non-transmissions in each family were stored for use by multi-locus analysis. A test for significance was provided with a permutation method creating a new data set with each pair of transmitted and untransmitted alleles, arbitrarily reversing the assignment of which was transmitted. The TDT was calculated for 10000 permutation repeats for single marker analysis and 1000 repeats for 2 marker analysis, comparing each simulated data set to the actual results observed in the real data set. In the data output it is indicated how many of the permuted data sets had a higher maximum value and how many had more results above certain thresholds (0.01, 0.001) of significance. In each program association was calculated for up to three-marker haplotypes

passing a sliding window through the marker map. Peaks of association were subject to further investigation in the candidate gene approach.

Beside the investigation of association in the SNP marker map, an analysis of linkage was performed using Genehunter, in order to delineate the non-parametric LOD score (NPL) in the marker set analysed (40). With this method a non-parametric trait like the disease affection status (affected or unaffected) can be analysed for linkage within a genomic region. The diallelic SNP markers were employed as genomic markers to which recombination was measured within the ASPs. To enlarge the number of ASPs in the analysis as well dependent ASPs were included. To avoid the bias the dependent sibpair was weighted less and added less linkage that the independent ASPs. Together with the linkage scan, the overall count of recombinants was calculated along with the value expected on the basis of the given physical distance between the SNP markers for the entire data set. If recombination was observed significantly more often than expected this could be an indication of a marker that was error-prone over multiple pedigrees or of an error in the entered genetic map, either in order or distance.

As a measure of inter-marker linkage disequilibrium (LD) the Lewontin's D' (149, 150) was calculated employing an analysis option of the HAPMAX program (http://www.uni-kiel.de/medinfo/mitarbeiter/krawczak/download/hapmax.exe). The measure estimates the difference between the number of two marker haplotypes observed and the one expected under the assumption of independence in segregation of markers using the correlation coefficient. Thereby the influence of allele frequency is considered in the calculation. Markers with a frequency below 1% in the control cohort were excluded from this analysis.

For the analysis of HLA-DPA1 the same statistical association methods were employed. Pearson's χ^2 statistics were calculated on contingency tables for alleles and genotypes as well as the frequencies for each polymorphic site and for shared epitopes over the three polymorphic sites (43). Randomly selected cases from the families were taken from the German and the South African family sample. From the South Korean population sample of unrelated individuals all cases available were included in the analysis. As control samples for the German population unrelated blood donors were used. Unrelated control samples were available as well from the South Korean population. In the South African population a randomly selected healthy family member was chosen as control sample from each family.

Transmission distortion was analysed in the German and South African family cohort using TRANSMIT (148) and the TDT using Genehunter (40). Inter-marker LD between the three polymorphic sites was calculated using EH (151), which is able to estimate LD between multiallelic markers.

For the candidate genes with significant initial association signals the case control and the TDT association were recalculated with all markers from the gene including newly generated ones from the mutation detection for each disease category. Furthermore the odds ratio (OR) was calculated for these SNP markers as a measure of increased risk of affection with a certain SNP variant. Stratification was performed for the NOD2 +/- genotypes in the IBD category according to allele status at the NOD2 gene mutations (3020insC, R702W, L1007P) (63-65). The SNP markers from the candidate genes were as well analysed for association with the following severity factors of IBD: (i) development of fistulae, (ii) formation of stenosis and (iii) previous resection of parts of the bowel, as described by the affected individuals in the questionnaire (appendix 8.4). If not annotated differently p-values were not corrected for multiple testing.

2.7.2. Analysis of linkage disequilibrium structure

In order to estimate the genetic differences between the 5 populations included into this study the F-statistics of population subdivision (F_{ST}) were calculated with the Arlequin program (152). The F_{ST} value is a measure for the deficiency of heterozygotes in the total of all populations that could be explained by subdivision into the original populations (153).

Pairwise LD between SNP marker was quantified using Lewontin's standardized deviation coefficient D' (149, 150) comparing estimated two marker haplotypes with expected haplotypes in each population with the HAPMAX program (http://www.uni-kiel.de/medinfo/mitarbeiter/krawczak/download/hapmax.exe). This was done for the large 10 Mb region in the US American whites sample as well as for the smaller region of 3.53 Mb with the higher marker density, and in all 5 populations. In these analyses, markers were ignored for populations in which the rare allele occurred at a frequency of less than 5%. For the German population as well an allele frequency cut-off of 20% was adopted in order to reduce the influence of more recently evolved SNPs upon the inferred haplotype patterns. Markers were clustered according to D' values using a standard UPGMA (Unweighted Pair

Group Method with Arithmetic Mean) algorithm (154). In this procedure the D' value for all marker pairs was compared and the two markers with the highest D' value were connected through a dendrogram. Afterwards the arithmetic mean was calculated for this pair and for the further clustering process used as new D' value. Then the next highest D' value is identified and the markers clustered. Through this process sometimes markers were combined not neighbouring in the physical map. The marker order was kept as close as possible to the physical order by minimizing the number of non-sequentially clustered markers through rotation of branches in the emerging dendrogram. A clustering of D' values in this way has not been described before and allows the identification of linkage structures without previous assumptions.

LD structures in the 3.53 Mb region were compared between the different populations, employing the German population as reference population. The method developed for this purpose was to calculate Pearson's correlation coefficients of pair-wise D’ values markers in a sliding 500 kb window. In order to control sample size effects, random subsamples of 45 individuals (i.e. the minimum population size available for study) were drawn from the Norwegian, and the UK population and all individuals from the two US populations. From the German population the random sample of 45 individuals was drawn 1000 times and the correlation coefficients averaged over all 1000 subsamples. In addition, 95% confidence intervals were estimated for the German population sample from this procedure.

Metric LD units

This analysis was performed through Francisco De La Vega at Applied Biosystems, Foster City, California 94404 / USA. Due to reasons of consistency it is not separated from the study.

Metric LD units (LDU) were calculated, using the LDMAP program available at http://cedar.genetics.soton.ac.uk/public_html. This method has been developed to measure LD in the context of physical position of recombination (155). The LDU are additive over the physical distance with an increase in LDU representing a site of recombination. One LDU corresponds to the so-called “swept-radius” which has been suggested as the maximum (physical or genetic) distance over which useful LD can be expected (156). It equals the inverse of the exponential LD decay constant ε. LD map construction depends on obtaining an estimate for ε_i for the i^{th} interval between adjacent pairs of SNPs, using the information in all

SNPs typically within 100 kb. Two-marker haplotype data from SNP analysis were directly used as input to the LDMAP program to fit the Malecot model (156) for the decline of association with distance with expected value. For visualization of the LD map, the cumulative sum $? _{i=1..n}\varepsilon_i d_i$, measuring the distance of the n-th marker from the map origin in LDUs, was plotted against physical distance $? _{i=1..n} d_i$. Here, d_i is the physical length (in kb) of the i^{th} marker interval (155).

3. Results

3.1. Association mapping results

3.1.1. Case-control association

In order to further confine the area previously defined through linkage analysis in which a gene causing IBD, CD or UC might be located, a map of SNP markers was established within a 22.8 Mb long region on the human chromosome 6p covering the MHC region. After the exclusion of markers not matching the quality criteria a total of 142 SNP markers were employed for the association mapping (Fig. 2.2). For the analysis of association in the case control study design, the population was classified into the diagnostic categories healthy and affected (IBD), and the affection categories CD and UC. Cases for each category were taken from the monoplex families plus one randomly selected case from each multiplex family. The SNP markers, their position according to NCBI built 32 and the allele and genotype frequencies for each category can be seen in Table 8.1 appendix 8.1.1. Pearson's χ^2 was calculated for allele and genotypes.

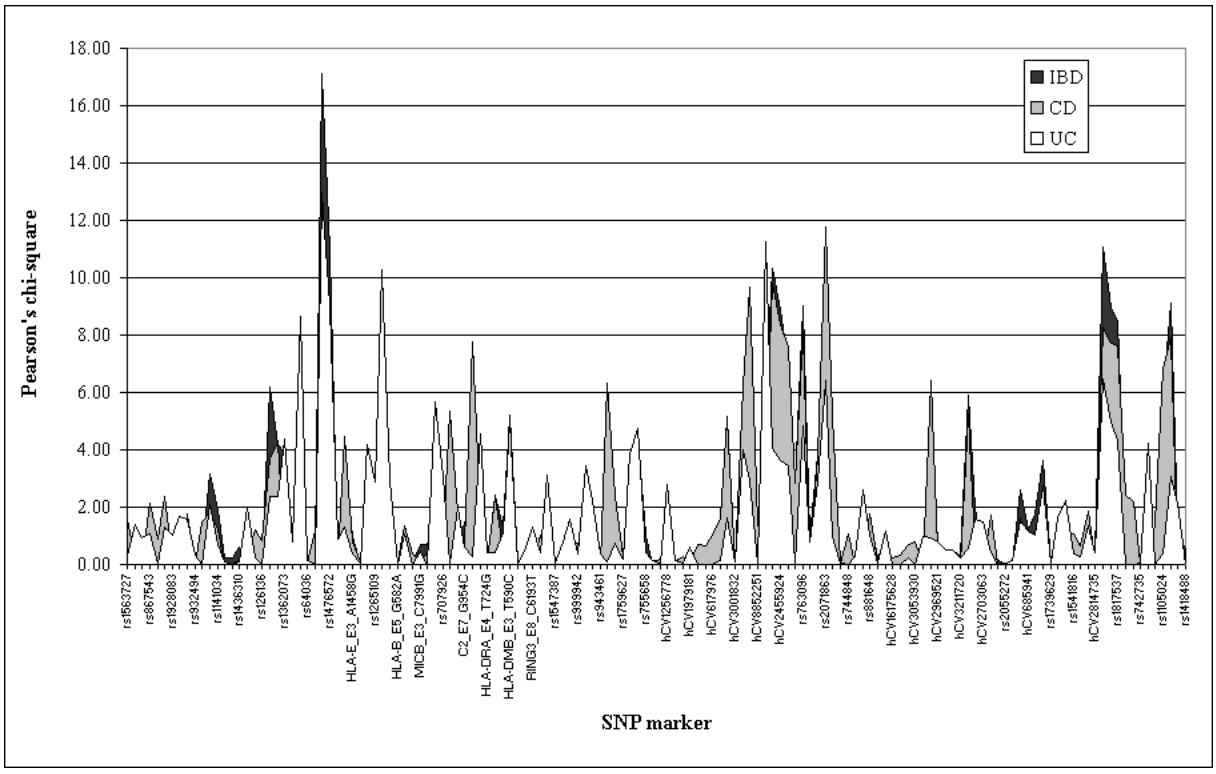

Fig. 3.1: Allele association analysis. Pearson's χ^2 was calculated for each of the 142 SNP makers. Not all markers are listed on the x-axis and the distance is not representing the physical distance. Significance of a peak at the 95% confidence interval is reached at a χ^2 value of 3.9 with 1 df.

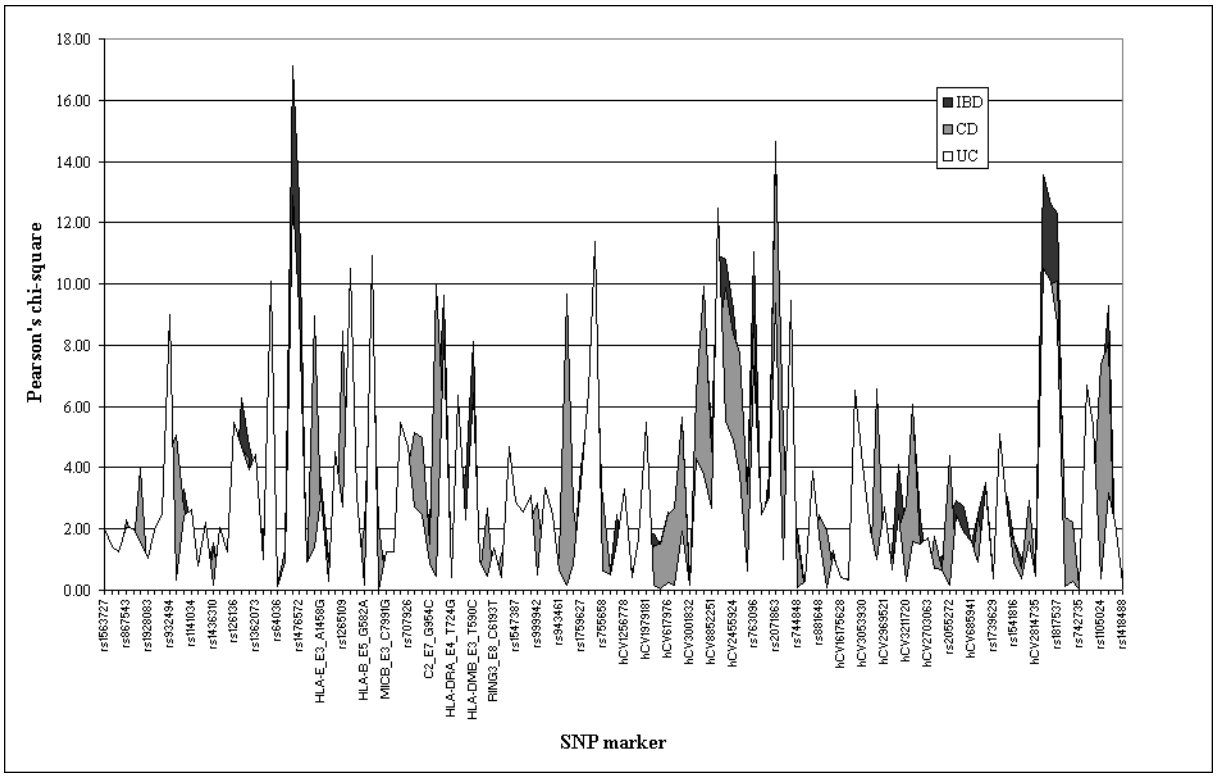

Fig. 3.2: Genotype association analysis. Pearson's χ^2 was calculated as described in Fig. 3.1. Significance of a peak at the 95% confidence interval is reached at a χ^2 value of 6.0 with 3 df.

The association for each marker in the allele and the genotype analysis is shown in Fig 3.1 and Fig. 3.2. Different peak associations can be identified for each disease category and for allele and genotype association. There were 8 peaks of association with IBD in the allele association analysis that reached a significance level of $p<0.05$. Peaks were characterised by an increased χ^2 value in several markers rising from both sides to the highest observed value in that physical region. Several markers and resulting peaks were mostly carried through one of the two IBD categories (CD or UC). For some markers, both CD and UC affected individuals contributed equally to the association. For CD 10 markers showed peak significant results in allele association, for UC 9 peak association markers were observed. The peak association markers for all categories are shown in Table 3.1, together with significance levels. It could be observed that markers within the same gene were associated to a different degree with the three categories of disease. From the makers with the highest χ^2, value several had an overall allele frequency below 10% (MOG_E3_G550A, rs902196, rs1059234, rs1105024, rs9395). This needs to be considered for the power of the association, and as well for the degree of inter-marker linkage. From the SNP marker map three areas (HSP70_Hom to HLA-G, SLC26A8 to MAPK13, and around TREM1) with major association leads were observed, each composed of several peaks with variable association intensity for the three disease categories. Several smaller peaks are located between those areas.

Gene	SNP marker name	c^2	p-value
IBD peak allele association			
	rs209179	6.17	0.01307
MOG	MOG_E3_G550A	17.10	0.00010
HLA-DMB	HLA-DMB_E3_T590C	5.23	0.02218
MAPK14	hCV2455926	10.33	0.00131
MAPK13	rs763096	9.04	0.00267
MAPK13	rs2071863	10.87	0.00098
TREM1	rs1817538	11.05	0.00089
SRF	rs9395	9.14	0.00253
CD peak allele association			
RFP	rs929042	4.21	0.04025
HCG-V	HCG-V_3pu_T3056C	4.48	0.03439
HSP70-HOM	HSP70-HOM_G2763A	5.37	0.02049
AGER	rs1035798	7.78	0.00534
	rs902196	6.34	0.01182
SLC26A8	hCV3001847	5.10	0.02387
MAPK14	hCV3001824	9.66	0.00190
CDKN1A	rs1059234	6.42	0.01132
PPIL1	hCV2061156	5.25	0.02195
PTK7	rs1105024	6.83	0.00901
UC peak allele association			
	rs1362073	4.35	0.03700
UBD	rs362536	8.64	0.00333
CDSN	CDSN_E1_A66T	4.20	0.04052
SC1	SC1_E2_G2712A	10.26	0.00136
DDAH2	rs707916	5.68	0.01725
BTNL-II	BTNL-II_E3_A158G	4.57	0.03262
TEAD3	hCV2489242	4.76	0.02913
MAPK14	hCV2455932	11.25	0.00080
CCND3	rs1410492	4.22	0.03999

Table 3.1: SNP markers indicating association peaks in the allele association.

In the genotype association analysis, the peak associations were partly shifted, but the main leads were still observed. Eleven markers showing peak associations were identified in IBD. Some of the SNPs with association to only one disease category in the allele association showed the highest χ^2 values for the combined category IBD in the genotype association. Eight peak association markers were observed for CD and 7 in UC (Table 3.2). Again low frequency was present in some SNPs. Frequencies of 10% or lower for the heterozygous state were observed for the following markers with peak association: rs902196, rs1105024 and rs9395. Overall Pearson's χ^2 values were higher than in the allele association analysis. Significance was not increased as in the analysis of genotypes a higher degree of freedom has to be taken into account (now 3 df instead of 1 df in the allele association). Even though the SNP markers with peak significance had shifted partially the areas in which association was highest were similar to the results from the allelic association.

Gene	SNP marker name	c^2	p-value
IBD peak genotype association			
	rs209179	6.29	0.04315
MOG	MOG_E3_G550A	17.12	0.00019
	rs2005423	9.92	0.00701
BTNL-II	BTNL-II_E3_A158G	9.65	0.00804
HLA-DMB	HLA-DMB_E3_T590C	8.14	0.01712
TEAD3	hCV2489242	9.48	0.00874
MAPK14	hCV2455932	10.94	0.00422
MAPK13	rs763096	11.06	0.00397
MAPK13	rs2071863	14.64	0.00066
TREM1	rs1817538	13.54	0.00115
SRF	rs9395	9.28	0.00966
CD peak genotype association			
HCG-V	HCG-V_3pu_T3056C	8.94	0.01143
	rs1265109	8.47	0.01446
AGER	rs1035798	9.98	0.00679
	rs902196	9.68	0.00790
MAPK14	hCV3001824	9.91	0.00704
CDKN1A	rs1059234	6.55	0.03776
PPIL1	hCV2061156	6.06	0.04831
PTK7	rs1105024	7.38	0.02500
UC peak genotype association			
	rs932494	8.99	0.01114
UBD	rs362536	10.11	0.00639
SC1	SC1_E2_G2712A	10.50	0.00526
HLA-DQA1	HLA-DQA1_E3_C389T	6.36	0.04152
	hCV2476745	9.47	0.00877
STK38	hCV791696	6.55	0.03787
CCND3	rs1410492	6.68	0.03536

Table 3.2: SNP markers indicating association peaks in the genotype association.

3.1.2. TDT association

Besides the case control study design the association to disease in the predefined linkage region was tested using family based transmission distortion. Two TDT algorithms were applied to assess association. In the Transmit program a comparison of the number of alleles transmitted from healthy parents to affected offspring with the number of alleles transmitted according to HWE was employed (148). In the Genehunter program the number of alleles transmitted from parent to offspring was compared to the number of alleles untransmitted (40). The significance of the association was calculated using χ^2 and the p-value at the 95% confidence interval. The association was calculated for up to three-marker haplotypes in a sliding window. For the single point association with Transmit the results are shown in Fig 3.3. Several association peaks were observed, again for the combined disease category IBD

and as well for the separated categories CD and UC. For IBD 7 peaks were identified, for CD another 7 markers show peak significant associations and for UC 4 markers were identified indicating the association peaks. The relevant markers for each disease category together with the number of alleles transmitted and the expected transmission are shown in Table 3.3.

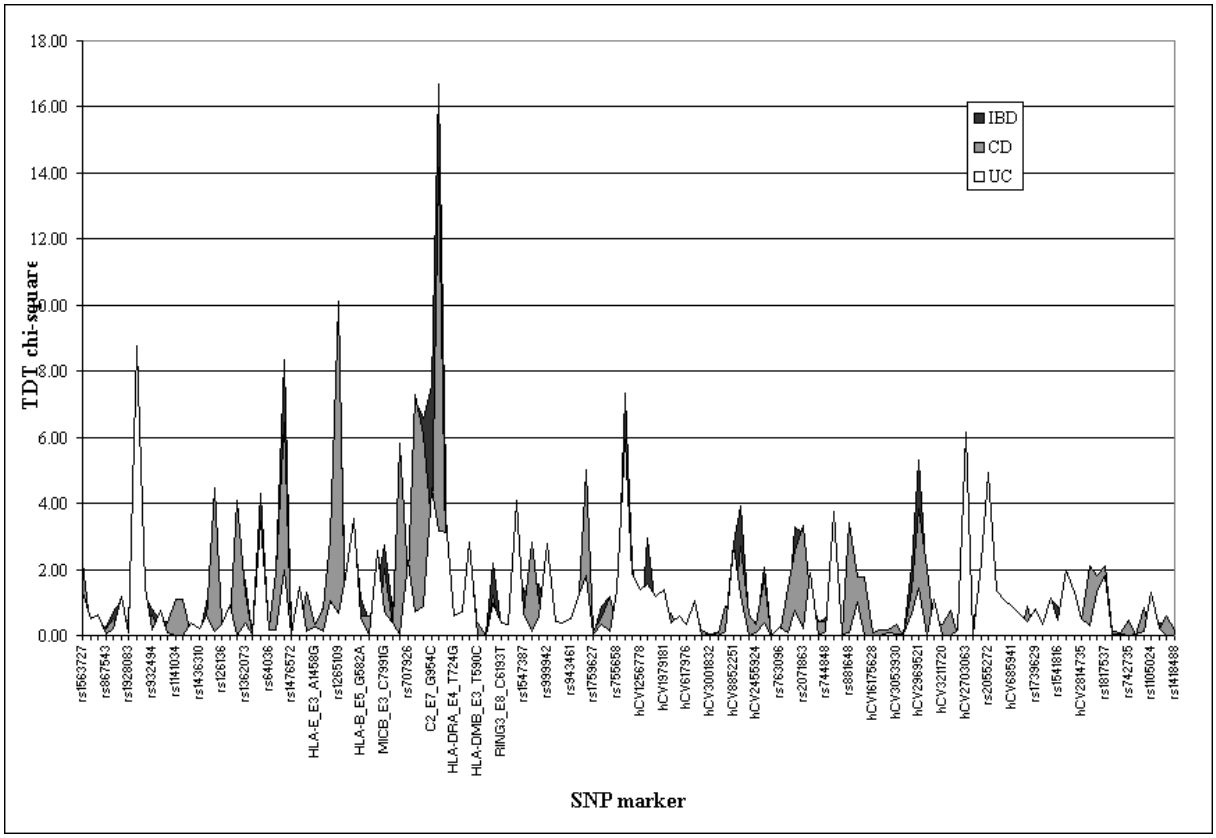

Fig. 3.3: Single point TDT (Transmit) association analysis. The χ^2 was calculated on the basis of the alleles observed to be transmitted compared to the number of alleles expected from the distribution in the parents according to HWE. Significance of a peak at the 95% confidence interval is reached at a χ^2 value of 3.9 with 1 df. Not all markers are shown and marker distance is not according to the physical distance.

As a measure for significance testing a bootstrap with 1000 repeats was performed. The p-values from the bootstrap analysis are shown in the Table 3.3. The bootstrap p-value should be reflecting the significance more appropriately when transmission is not uncertain. Significance of the bootstrap p-value was lower and some association peaks became insignificant: hCV2455932 and rs929042, or borderline significant: rs362536, rs126007. The most prominent lead was partly overlapping with the association described in the case control study (HSP70-Hom to C2). Several smaller association peaks near the two other main association leads from the case control analysis were observed but had little or no overlap.

Table 3.3: SNP markers indicating association peaks in the single point TDT

Gene	SNP marker name	c^2	p-value	Bootstrap p-value	Transmission ratio (observed/expected)
IBD TDT Transmit single marker association					
UBD	rs362536	4.31	0.03794	0.046	836/808.69
MOG	MOG_E3_G550A	8.36	0.00384	0.008	1405/1428.2
C2	C2_E7_G954C	7.56	0.00596	0.009	1567/1552
AGER	rs1035798	16.69	0.00004	0.001	1166/1118.5
FKBP5	rs992105	7.35	0.00669	0.021	1336/1362.6
MAPK14	hCV2455932	3.94	0.04721	0.055	1640/1623.5
CDKNA1	hCV2969521	5.32	0.02105	0.019	988/1020.2
CD TDT Transmit single marker association					
	rs126007	4.48	0.03439	0.046	808/827.24
RFP	rs929042	4.07	0.04359	0.051	842/823.73
	rs1265109	10.11	0.00147	0.001	900/868.48
DDAH2	rs707916	5.80	0.01601	0.031	638/663.52
HSP70-HOM	HSP70-HOM_G2763A	7.29	0.00693	0.006	754/726.54
HSPA1B	rs539689	6.08	0.01370	0.016	567/540.48
	rs914815	5.02	0.02511	0.032	577/602.01
UC TDT Transmit single marker association					
	rs1325016	8.77	0.00306	< 0.001	499/479.75
RXRB	rs926424	4.09	0.04308	0.022	354/368.5
MTCH1	hCV2703063	6.15	0.01315	0.017	314/335.29
FLJ20337	rs2055272	4.95	0.02615	0.029	345/361.31

In the two-marker analysis short haplotypes were estimated from the alleles of 2 neighbouring SNP markers. While the pattern of the plotted two point TDT χ^2 (Fig 8.1 appendix 8.1.2) exhibited a similar pattern as in the single marker analysis, the peak significant association marker-haplotypes were sometimes slightly shifted. The results were similar for the 3-marker haplotype analysis. The association peaks were overlapping with the two-marker peaks (Table 3.4). The frequency of several 3 marker haplotypes were very small, therefore the differences between expected and observed transmission was overestimated causing a spuriously significant result.

Table 3.4: SNP markers indicating association peaks in the two point TDT analysis.

Gene	SNP marker name	c^2	p-value	Bootstrap p-value	Transmission ratio* (observed/expected)	Frequency*
IBD TDT Transmit 2 marker association						
RFP	rs929042	9.20	0.02675	0.037	301.72/325.1	0.18
MOG	MOG_E3_G550A	10.65	0.01377	0.008	200.24/175.86	0.08
C2	C2_E7_G954C	26.62	0.00001	0.000	1174.7/1116.3	0.65
RXRB	rs926424	9.06	0.02851	0.019	899.83/929.11	0.56
FKBP5	rs755658	9.72	0.02107	0.031	283.4/254.31	0.16
MAPK13	rs2007683	11.25	0.01045	0.004	2.7878/6.1433	0.07
CDKNA1	hCV2969521	12.89	0.00487	0.007	776.9/819.49	0.39
CD TDT Transmit 2 marker association						
GABABR1	GABABR1_E7_G141esA	14.53	0.00227	0.005	125.71/107.47	0.11
CDSN	CDSN_E1_A66T	10.92	0.01214	0.012	259.71/289.9	0.20
VARS2	rs707926	8.82	0.03172	0.030	309.23/339.48	0.29
	rs914815	10.24	0.01665	0.018	474.96/444.87	0.38
TREM1	rs1351835	9.02	0.02904	0.012	0.52107/2.0837	0.13
UC TDT Transmit 2 marker association						
	rs1928083	10.26	0.01647	0.012	41.66/52.038	0.09
	rs1325016	11.74	0.00833	0.012	47.53/63.705	0.13
	rs302976	8.40	0.03843	0.029	101.73/91.25	0.14
SRPK1	rs1326752	8.62	0.03485	0.000	0.77853/3.1301	0.13

*of the significant haplotype

The plotted TDT association had a similar picture for the single marker association in Genehunter (Fig 3.4) than in the Transmit analysis event though the peaks were more prominent in the Transmit analysis. The TDT single marker analysis with Genehunter identified fewer significant association leads than the Transmit analysis (Table 3.5). The association peaks that were observed were overlapping with those from the Transmit analysis. Differences were only seen for the SNP in the gene AGER that was associated with IBD in the Transmit analysis and with CD in the Genehunter analysis, and for the SNP in HLA-C, that did not reach significance in the Transmit analysis. In the Genehunter program a permutation test to determine TDT significance was performed. For the single marker analysis of IBD 2308 out of 10000 permutations tested had a larger maximum value than the real best (9.72). For CD 1821 of 10000 had a larger maximum value than the real best (10.20) and for UC the highest value observed was 11.92 which was exceeded in 639 of 10000 permutation tests.

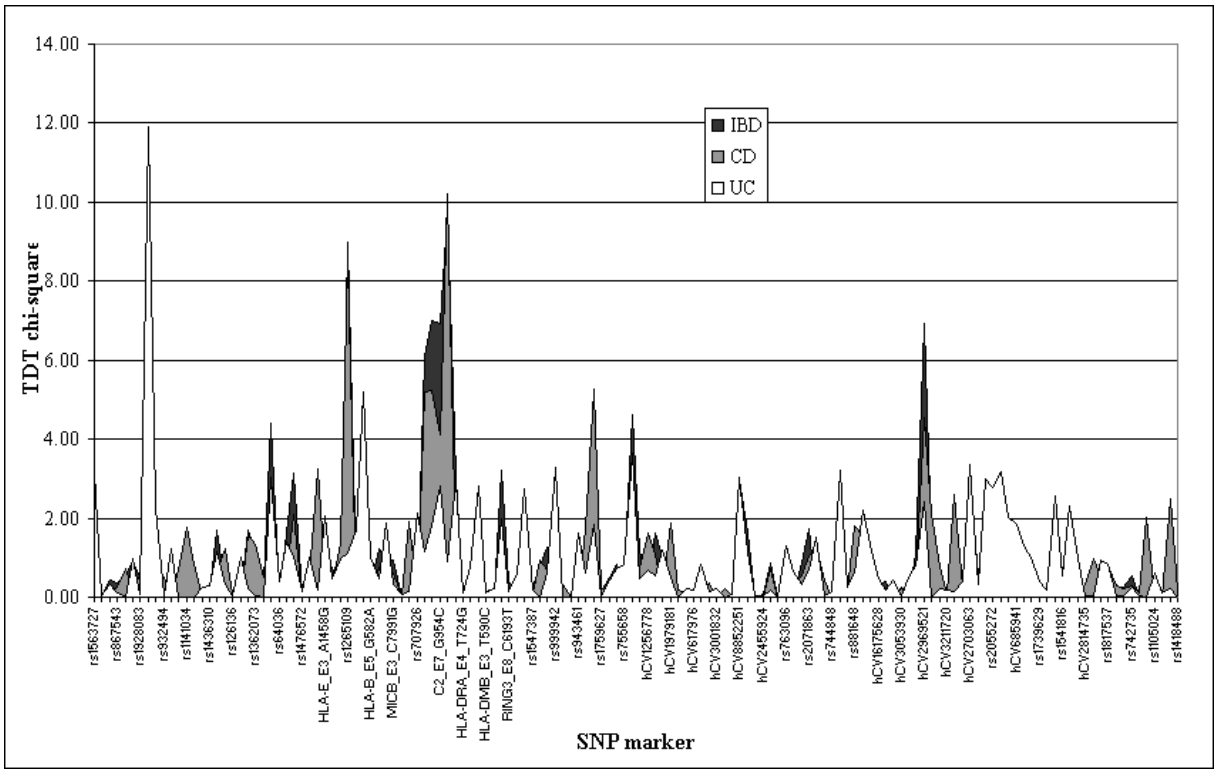

Fig. 3.4: Single point TDT (Genehunter) association analysis. The χ^2 was calculated on the basis of the alleles transmitted from parent to affected offspring compared to the not transmitted alleles. Significance of a peak at the 95% confidence interval is reached at a χ^2 value of 3.9 with 1 df. Not all markers are shown and the positioning is not according to the physical distance.

Table 3.5: SNP markers indicating association peaks in the Genehunter single point analysis.

Gene	SNP marker name	c^2	p-value	Times transmitted	Times untransmitted
IBD TDT Genehunter single marker association					
UBD	rs362536	4.39	0.03616	334	282
FKBP5	rs992105	4.63	0.03145	137	175
CDKNA1	hCV2969521	6.94	0.00842	367	299
CD TDT Genehunter single marker association					
	rs1265109	8.52	0.00351	150	205
AGER	rs1035798	10.20	0.00141	170	116
	rs914815	5.27	0.02167	186	233
UC TDT Genehunter single marker association					
	rs1325016	11.92	0.00056	50	91
HLA-C_	HLA-C_E2_G1019A	5.21	0.02245	88	121

In the 2 marker analysis a number of significant association marker haplotypes for IBD were found as well in the Genehunter analysis: rs362536 - rs64036, CDSN_E1_A66T - rs1265109, HSP70-HOM_G2763A - rs539689, hCV2969521 - hCV1022631. The highest association was observed for C2_E7_G954C - rs1035798 (16.64, p-value 0.000045). The rate of transmitted/untransmitted was 227/148 for the haplotype carrying the greatest significance in the C2_E7_G954C - rs1035798 marker combination. Significant results that were caused by very low numbers of haplotypes transmitted and untransmitted are not listed here. The

permutation analysis identified 25 out of 1000 permutations tested with a higher maximum value then the real best. A similarly high value was not observed in the CD category and the permutation analysis identified 70 of 1000 testes with a larger maximum value than the real best (14.44 C2_E7_G954C - rs1035798, p-value 0.000144). For the UC category the highest association observed was 11.44 (HLA-C_E2_G1019A - HLA-B_E5_G582, p-value 0.000721), the permutation showed 296 of 1000 tests with a larger maximum value than the real best. The situation for the 3 marker analysis was similar with peak associations in the same regions. For the disease categories CD and UC no better association results were observed.

3.1.3. Additional analyses

A non-parametric linkage (NPL) analysis was performed to verify the presence of linkage in the region analysed and to evaluate the intensity of linkage to the disease trait. The 142 SNP markers from the association study were analysed using their physical position as determined from NCBI build 32. The NPL for IBD is shown in Fig 3.5, the plots for CD and UC are in the Appendix 8.1.3. (Figs 8.2, 8.3) The analysis was performed using Genehunter. Nonparametric linkage is present in the region on chromosome 6p analysed for IBD, for CD alone the NPL score is decreased, and for UC alone the NPL score was very low over the complete region. For the disease category IBD the NPL score reached values of 3.47 at two positions and was above 3 over a large region. To the centromeric side the NPL score decreases to 0.5 in the last position clearly marking the centromeric end of the linkage region. This decrease was not observed on the telomeric side of the region. Maximal values for CD were at 2.6 and for UC all values were below 1. The curve for CD is similar to the IBD NPL score but on a lower level, while the UC analysis is not comparable. The NPL is calculated from affected sibling pairs and strongly dependent on the population size. In the category IBD 181 ASPs were analysed. For CD the number of ASPs in the analysis was 105 and for UC only 43 ASPs were available for the analysis. This could explain some of the difference in the NPL score. In the detailed analysis of the linkage region on chromosome 6p 428 ASPs were involved

Recombination was calculated as part of the linkage analysis and compared to the expected recombination value (theta) between each marker. The only recombination values observed to

be higher than expected were between the marker rs747694 and hCV2489242, however not significantly.

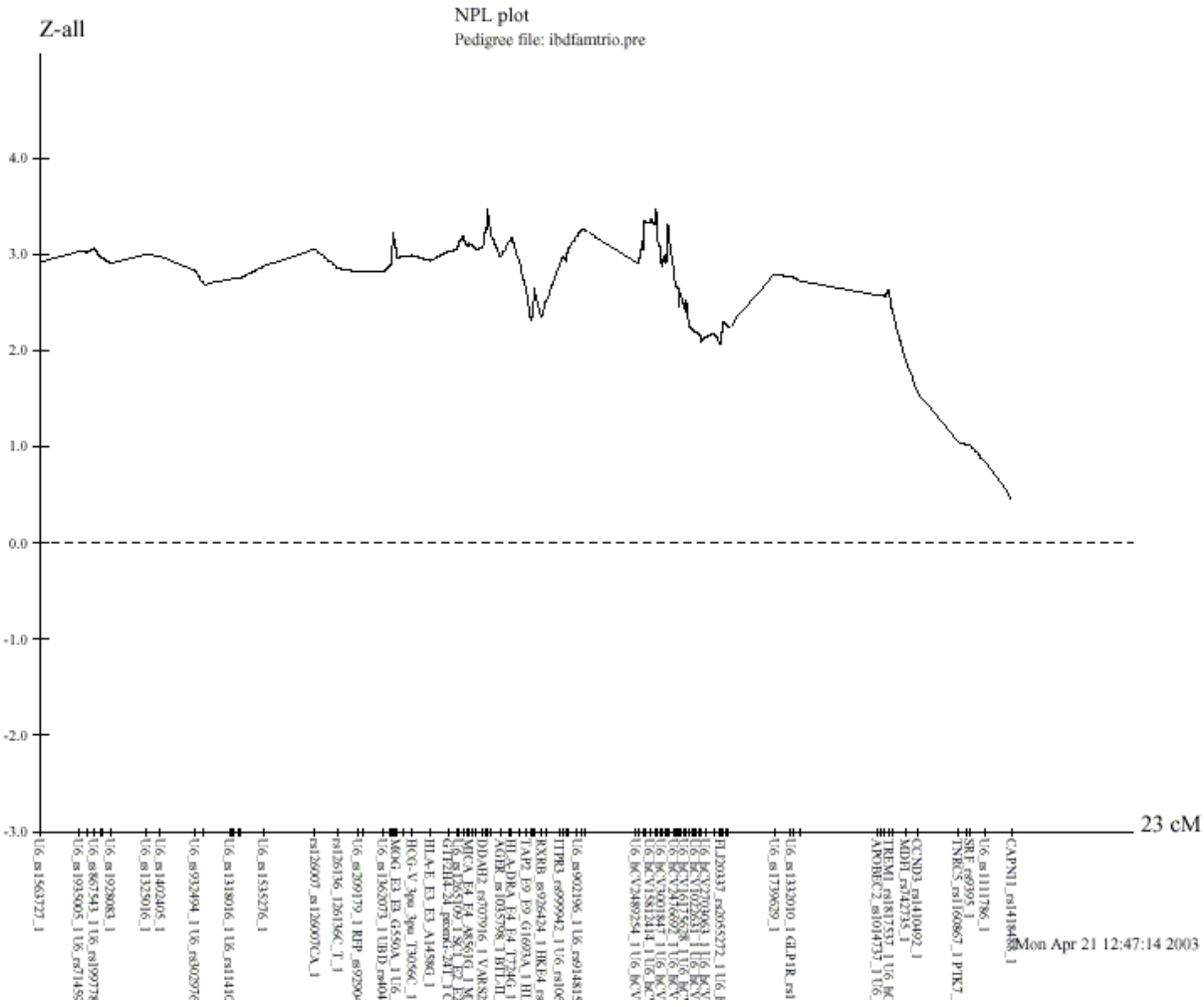

Fig. 3.5: Non-parametric LOD score for IBD. The NPL was calculated on the basis of 208 ASPs (181 independent ASPs) using the weighed pairs option of Genehunter. The NPL score is indicated as Z-all on the y-axis. The x-axis shows all markers involved and their physical position as short black lines. Marker names could not be indicated completely All NPL values were highly significant ($p < 0.001$)

In order to reveal the LD between SNP markers Lewontin's standardized deviation coefficient D' was calculated (149, 150). Certain areas with a high internal D' value (< 0.7) could be observed in the SNP map. Mostly these included SNP marker from one gene or physically close positioned genes, but some markers exhibited a high LD over long distances. Blocks with high LD were formed by the markers covering SLC26A8 to MAPK13, TEL2, TREM1 and FKBP5. Regions with a generally high D' value but less clear boundaries were observed in the region around TNF-α and UBD and the telomeric side of the SNP map. One of the UBD marker exhibited long range LD as did rs64036, the MOG marker, rs902196 and

rs763017. Some of these markers showed significant association with IBD CD or UC in the case control or the TDT analysis.

3.2. Candidate genes

3.2.1. HLA-DPA

Three polymorphic sites in the HLA-DPA1 gene at nucleotide positions 91/92, 111/ 114 and 149 in exon two were genotyped in a German, a South African and a South Korean population. The polymorphic structure of the region on exon 2 of HLA-PDA1 with the resulting amino-acid exchanges and the corresponding HLA-DPA1 alleles are shown in Table 3.6.

Table 3.6: HLA-DPA1 shared epitopes, nucleotide exchanges and corresponding amino acids (AA) and alleles.

Shared epitope	Nucleotide 91,92	AA 31	Nucleotide 111, 114	AA 37-38	Nucleotide 149	AA 50	Corresponding Allele DPA1*
1	GAG**AT**GTTC	Q	CTGGA**C**AA**G**AAG	D-K	GCC**A**AGCC	Q	0103; 0104; 0105; 0301; 0302
2	GAG**CA**GTTC	M	CTGGA**T**AA**A**AAG	D-K	GCC**G**AGCC	R	02011; 02012
3	GAG**CA**GTTC	M	CTGGA**C**AA**G**AAG	D-K	GCC**G**AGCC	R	02013; 02022
4	GAG**AT**GTTC	Q	CTGGA**C**AA**G**AAG	D-K	GCC**G**AGCC	R	0203; 0401
5	GAG**CA**GTTC	M	CTGGA**T**AA**A**AAG	D-K	GCC**A**AGCC	Q	106
6	GAG**CA**GTTC	M	CTGGA**T**AA**G**AAG	D-K	GCC**G**AGCC	R	2021

Allele and genotype frequencies in the German and the South-Korean cohorts of UC, CD and IBD cases did not differ significantly from the frequency in healthy controls of the same ethnic background. An overview of the allele and genotype frequencies is given in Table 3.7 and Table 3.8. Only the analysis of the South African population indicated a trend towards a higher allele and genotype frequency of the sequence GATAAA at the position 111/114 in UC patients in comparison with controls (allele frequencies: χ^2 = 8.66, p = .0131, 2 df; genotype frequencies: χ^2 = 12.7, p = 0.0127, 5 df). However, the South African population comprised 20 affected individuals in the UC category (Table 2.3) and a similar result was not observed in any of the other populations.

Suggestive evidence for TDT association in the German population sample was observed for CD and the HLA-DPA1*02021 / *02016 allele, corresponding to shared epitope type 6 (Table

3.6). The corresponding TDT-χ^2 for linkage at this position was 4.5 with a nominal p-value of 0.043. No further association was seen in the German or the South African family sample and the South Korean case control sample (all nominal p-values >0.1).

Table 3.7: HLA-DPA1 exon 2 polymorphism allele frequencies in three different populations.

Nucleotide position		European family population				South African family population				South Korean population			
		Cases			Controls	Cases			Controls	Cases			Controls
		IBD	CD	UC	unrelated	IBD	CD	UC	family	IBD	CD	UC	unrelated
91, 92	C-A	0.16	0.18	0.14	0.18	0.32	0.28	0.45	0.36	0.55	0.52	0.58	0.52
	A-T	0.84	0.82	0.86	0.82	0.68	0.72	0.55	0.64	0.45	0.48	0.42	0.48
111, 114	C..G	0.86	0.84	0.87	0.83	0.78	0.84	0.50	0.80	0.84	0.89	0.82	0.89
	T..A	0.13	0.14	0.12	0.15	0.21	0.14	0.50	0.18	0.16	0.11	0.18	0.08
	T..G	0.01	0.02	0.00	0.02	0.01	0.02	0.00	0.02	0.00	0.00	0.00	0.02
149	A	0.85	0.83	0.87	0.82	0.66	0.69	0.50	0.67	0.42	0.41	0.40	0.46
	G	0.15	0.17	0.13	0.18	0.34	0.31	0.50	0.33	0.58	0.59	0.60	0.54

Table 3.8: HLA-DPA1 exon 2 polymorphism genotype frequencies in three different populations.

Nucleotide position		European family population				South African family population				South Korean population			
		Cases			Controls	Cases			Controls	Cases			Controls
		IBD	CD	UC	unrelated	IBD	CD	UC	family	IBD	CD	UC	unrelated
91, 92	C-A/A-T	0.26	0.28	0.23	0.33	0.43	0.28	0.70	0.49	0.37	0.43	0.34	0.45
	C-A/C-A	0.03	0.04	0.02	0.02	0.11	0.14	0.10	0.11	0.37	0.30	0.41	0.30
	A-T/A-T	0.71	0.68	0.75	0.65	0.47	0.59	0.20	0.40	0.26	0.26	0.25	0.25
111, 114	C..G/T..A	0.21	0.22	0.20	0.26	0.36	0.24	0.80	0.33	0.23	0.13	0.26	0.14
	C..G/T..G	0.02	0.04	0.01	0.03	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.04
	C..G/C..G	0.74	0.71	0.77	0.68	0.60	0.72	0.10	0.62	0.72	0.83	0.69	0.80
	T..A/T..A	0.02	0.03	0.02	0.02	0.02	0.00	0.10	0.00	0.05	0.04	0.05	0.01
	T..A/T..G	0.00	0.01	0.00	0.00	0.02	0.03	0.00	0.02	0.00	0.00	0.00	0.00
149	A/G	0.25	0.25	0.22	0.33	0.47	0.34	0.80	0.49	0.40	0.48	0.38	0.42
	A/A	0.72	0.71	0.76	0.66	0.43	0.52	0.10	0.42	0.22	0.17	0.21	0.25
	G/G	0.03	0.04	0.02	0.02	0.11	0.14	0.10	0.09	0.38	0.35	0.41	0.32

Linkage between the three polymorphic sites has been estimated using EH. They were in strong LD (pairwise χ^2 > 800.00, p < 0.001, 2 df), and the haplotypes of the six shared epitopes described in Table 3.6 were highly significant (χ^2 = 1947.51, p < 0.0001, df = 11). Analysis of the three most relevant shared epitopes showed a trend towards a higher frequency of the shared epitope type 2 containing the sequence GATAAA at position 111, 114 (χ^2 = 5.66, p = 0.058, df = 3) in the South African population (HLA-DPA1*02011/02012/02014) (Table 3.9).

No differences were seen in the German and South Korean populations. Frequency distribution of the shared epitopes compared to the allele frequency in the three populations showed, that they were within the range for the ethnicity (157).

Table 3.9: HLA-DPA1 exon 2 shared epitope frequencies in three different populations.

Shared epitope	German family population				South African family population				South Korean population			
	Cases			Controls	Cases			Controls	Cases			Controls
	IBD	CD	UC		IBD	CD	UC		IBD	CD	UC	
1	0.83	0.82	0.85	0.81	0.69	0.76	0.57	0.70	0.43	0.43	0.40	0.46
2	0.13	0.13	0.13	0.14	0.22	0.12	0.36	0.18	0.16	0.11	0.18	0.08
3	0.02	0.02	0.02	0.02	0.08	0.12	0.04	0.11	0.39	0.41	0.40	0.42
4	0.00	0.01	0.00	0.01	0.01	0.00	0.04	0.01	0.02	0.04	0.02	0.01
5	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
6	0.01	0.02	0.01	0.02	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02

3.2.2. Other candidate genes

The complete cDNA of the three splice variants of the MAPK14 (p38-α, CSBP1 and Mxi2) includes 13 exons. All exons and about 1000 nucleotides in the 5 prime regions were sequenced in 30 individuals, some exons have been sequenced in an additional 47 IBD samples. No polymorphisms in the exons or in the potential promotor region were found.

This is in contrast to data reported at the NCBI locus link SNP information for MAPK14 LocusID: 1432 (http://www.ncbi.nlm.nih.gov/LocusLink/), where several polymorphisms were reported for exon 9. Alignment of the three splice variants identified the exons 9 and 10 to be of identical length with a high degree of similarity. The presence of both transcripts in an analysis of cDNA or mRNA could be causing the annotation of SNPs. Mutation detection in MAPK13 and BRPF3 did not identify any SNPs in the exons, predicted exons of BRPF3 or potential promotor region of MAPK13. The sequence analysis of TREM1 exons revealed SNPs in the exon 2 and one in the 3' untranslated region (3'UTR). These SNPs in the meantime received an entry in the dbSNP with the identification numbers rs2234237 (A to T, A=1, T=2) and rs2234246 (A to G, A=1, G=2) (Fig.3.6). The exon 2 polymorphism results in an exchange of amino acids from T to S at the AA position 25. One more SNP (a G to A exchange, G=1, A=2) in the potential promotor region 217 bp before the start codon has not been described yet. The frequency of these three new SNPs is shown in Table 3.10. The genotype frequencies differ between the control individuals and the IBD cases by 7% in the homozygous 2-2 (G-G) variant for the rs2234246. For both other SNPs the greatest difference

was seen between the frequency in controls and UC cases, while in the IBD and UC category the frequency was more similar to the control individuals.

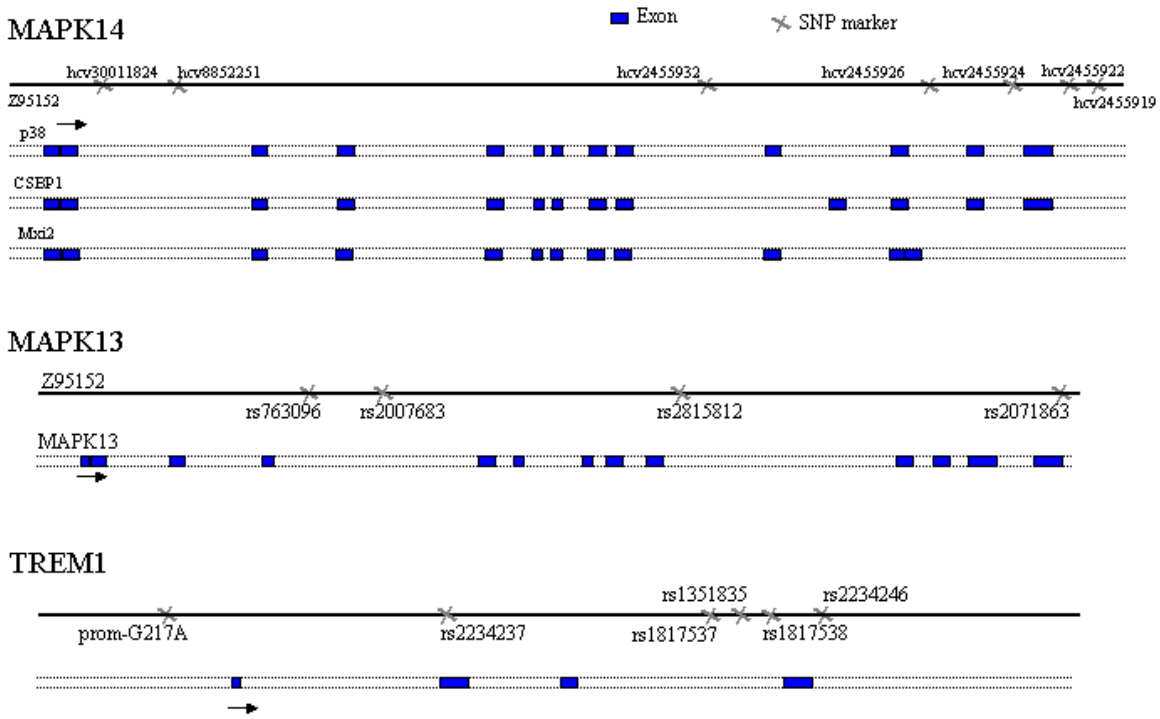

Fig. 3.6: Schematic description of the candidate genes MAPK14, MAPK13 and TREM1. Alternative splice versions are indicated as far as considered in this study. The candidate gene BRPF3 is not depicted as no polymorphisms were described for the gene and none could be identified during mutation detection.

Table 3.10: Allele and genotype frequencies in the new polymorphisms of TREM1.

TREM1		Allele frequency		Genotype frequency		
	Population	1	2	1-1	1-2	2-2
rs2234246	controls	0.50	0.50	0.27	0.46	0.27
	IBD	0.45	0.55	0.21	0.49	0.30
	CD	0.45	0.55	0.21	0.48	0.32
	UC	0.46	0.54	0.19	0.53	0.27
rs2234237	controls	0.90	0.10	0.82	0.17	0.01
	IBD	0.91	0.09	0.83	0.16	0.01
	CD	0.90	0.10	0.82	0.18	0.01
	UC	0.94	0.06	0.88	0.12	0.00
prom-G217A	controls	0.93	0.07	0.87	0.13	0.00
	IBD	0.90	0.10	0.82	0.16	0.02
	CD	0.90	0.10	0.82	0.16	0.02
	UC	0.91	0.09	0.83	0.15	0.01

Association for the new markers in the case control study design was weak but present for the IBD and CD disease category. The Pearson's χ^2 and corresponding p-values for all new SNPs

are shown in Table 3.11. Significance at the 95% confidence interval was reached only for the prom-G217A polymorphism in the allele and genotype analysis of the IBD category and in the allele association analysis of the same SNP in the CD category. The odds ratio (OR) was calculated for all SNP markers in the candidate genes. Generally the OR for the homozygous state was higher than for the heterozygous state.

Table 3.11: Allele and genotype association in the new polymorphisms of TREM1.

TREM1		Allele association		Genotype association	
SNP marker name		c^2	p-value	c^2	p-value
IBD	rs2234246	4.25	0.03923	5.82	0.05447
	rs2234237	0.76	0.38791	1.24	0.53844
	prom-G217A	5.92	0.01501	6.82	0.03302
CD	rs2234246	4.17	0.04115	4.48	0.10617
	rs2234237	0.03	0.87017	0.28	0.86858
	prom-G217A	4.37	0.03672	5.98	0.05022
UC	rs2234246	2.26	0.13489	5.67	0.05878
	rs2234237	3.51	0.06116	3.42	0.18058
	prom-G217A	2.77	0.09631	4.28	0.11723

For IBD the marker rs2234237 from TREM1 had the highest odds ratio with 1.96 however this marker had a 95% confidence interval of 0.59 - 7.30. From MAPK14 hCV2455919 was the marker with the highest OR (95% confidence interval 1.03 - 3.10). All values for CD were below that. For UC the marker rs2234237 exhibits an OR of 3.15 (95% confidence interval 0.65 - 4.95). The OR is affected through a low frequency of a marker, which shows in the large confidence intervals. The transmission distortion association for all three new markers did not show any significant results for the IBD disease category in the Transmit and the Genehunter analysis. For the subcategories CD and UC the results were similar, without significant results.

Stratification was performed for all markers from the candidate genes MAPK14, MAPK13 and TREM1 for the IBD category according to allele status at the NOD2 gene mutations (3020insC, R702W, L1007P) (63-65). After stratification for carriers of one of the three NOD2 mutations 72 control and 340 case individuals were analysed using the case control association. The stratification against the NOD2 mutations includes 332 controls and 482 cases. Mostly the association with IBD is reduced for both, the NOD2 carrier and the non-NOD2 carrier, which could be due to a reduced sample size. For the MAPK13 marker rs2071863 and the TREM1 marker rs1817538, rs1351835, and rs1817537 the association seen in the unstratified analysis seems to be carried mostly through the non-NOD2 carriers (Table

3.12). For all other SNPs the association seems to be contributed equally from both NOD2 mutation carriers and non-carriers.

Table 3.12: Allele and genotype association after stratification +/- NOD2.

SNP marker name	NOD2+				NOD2-			
	Allele association		Genotype association		Allele association		Genotype association	
MAPK14	c^2	**p-value**	c^2	**p-value**	c^2	**p-value**	c^2	**p-value**
hCV3001824	6.10	0.01356	9.73	0.00771	2.59	0.10801	4.51	0.10485
hCV8852251	0.18	0.67581	0.45	0.79961	0.24	0.63148	3.96	0.13788
hCV2455932	2.10	0.15093	4.30	0.11607	3.07	0.08036	7.21	0.02721
hCV2455926	5.17	0.02294	7.33	0.02564	3.29	0.07016	5.04	0.08030
hCV2455924	5.15	0.02317	7.76	0.02068	2.64	0.10513	4.09	0.12902
hCV2455922	1.05	0.30534	1.15	0.56311	4.37	0.03655	4.94	0.08440
hCV2455919	0.00	0.95487	0.11	0.94476	1.13	0.28761	2.63	0.26925
MAPK13								
rs763096	4.65	0.03106	5.35	0.06880	1.29	0.25866	3.88	0.14342
rs2007683	0.01	0.94533	0.11	0.94732	0.08	0.79992	1.24	0.53847
rs2815812	3.04	0.08154	4.06	0.13131	2.55	0.11117	3.70	0.15735
rs2071863	2.27	0.13452	2.34	0.31026	6.82	0.00906	11.76	0.00280
TREM1								
rs2234246	0.42	0.51962	0.48	0.78633	3.90	0.04819	7.34	0.02547
rs1817538	1.44	0.23418	1.46	0.48384	8.56	0.00349	12.89	0.00159
rs1351835	0.31	0.59138	0.30	0.86061	7.68	0.00566	12.65	0.00179
rs1817537	0.30	0.59733	0.30	0.86124	7.25	0.00716	10.82	0.00447
rs2234237	0.13	0.72622	0.89	0.64105	0.60	0.44361	1.21	0.54716
prom-G217A	0.00	0.96292	1.19	0.55221	3.80	0.05132	3.84	0.14635

The analysis of the TDT in the stratified samples showed only minor differences to the unstratified data. Significant association was not reached in any of the markers neither in the stratification for nor against NOD2 carriers. The data are only shown for the Transmit results in Table 3.13. In the TDT analysis using the Genehunter program produced similar or even weaker association.

Table 3.13: TDT association (Transmit) after stratification +/- NOD2.

SNP marker name	c^2	p-value	Bootstrap p-value	Transmission ratio* (observed / expected)	Freq.*	c^2	p-value	Bootstrap p-value	Transmission ratio* (observed / expected)	Freq.*
MAPK14	**stratified for CARD15 carrier**					**stratified for non-CARD15 carrier**				
hCV3001824	0.04	0.8475	0.847	438/436.16	0.54	0.16	0.6922	0.716	598/593.63	0.54
hCV8852251	2.92	0.0873	0.092	681/692.84	0.83	0.87	0.3498	0.327	942/949.43	0.83
hCV2455932	2.63	0.1047	0.089	743/734	0.88	1.50	0.2213	0.243	1014/1006	0.88
hCV2455926	0.06	0.8126	0.823	426/423.74	0.53	0.28	0.5965	0.626	590/584.11	0.53
hCV2455924	0.00	0.9977	0.997	434/433.97	0.53	0.02	0.8853	0.882	592/590.4	0.53
hCV2455922	0.16	0.6849	0.647	581/577.4	0.67	3.02	0.0824	0.091	747/764.92	0.67
hCV2455919	1.22	0.2701	0.313	669/677.24	0.81	0.58	0.4479	0.467	928/921.6	0.81
MAPK13										
rs763096	0.76	0.3826	0.377	427/435.04	0.60	0.15	0.7024	0.695	613/609.1	0.60
rs2007683	0.47	0.4928	0.483	472/466.55	0.65	1.45	0.2288	0.231	621/632.68	0.65
rs2815812	0.41	0.5245	0.509	582/576.66	0.72	2.01	0.1559	0.149	807/792.72	0.72
rs2071863	0.44	0.5056	0.500	676/671.09	0.81	1.89	0.1691	0.155	930/918.09	0.81
TREM1										
rs2234246	0.82	0.3661	0.387	401/392.77	0.54	0.69	0.4045	0.432	542/550.67	0.54
rs1817538	1.37	0.2420	0.269	405/394.22	0.54	0.07	0.7912	0.813	534/536.75	0.54
rs1351835	0.67	0.4135	0.412	395/387.66	0.53	0.31	0.5786	0.646	534/539.75	0.53
rs1817537	0.44	0.5050	0.528	401/394.85	0.54	0.21	0.6435	0.691	541/545.87	0.54
rs2234237	1.63	0.2024	0.196	678/684.37	0.89	0.48	0.4877	0.494	920/924.13	0.89
prom-G217A	0.16	0.6870	0.734	679/676.8	0.88	0.04	0.8395	0.839	913/911.7	0.88

*of the significant haplotype

Association of the markers in the candidate genes was analysed as well for three aspects considered to be markers of severity in the progression of IBD. The information was obtained from the questionnaire filled by the participants of the family based study (appendix 8.4). These aspects were (i) the development of fistulae, (ii) the formation of stenosis and (iii) a previous resection of parts of the bowel. Some affected individuals had two or all three aspects in their disease history, and therefore were included in all three analyses. For the association analysis of fistulizing disease 266 affected individuals were included into the calculation of Pearson's case control χ^2. Despite the reduced population sample several markers from MAPK14 showed slightly increased association with the fistulizing disease, while the association for the MAPK13 markers did not change and in the SNPs of the TREM1 gene no significant association was observed in contrast to the association with the disease category IBD (Table 3.14). The significant association in the MAPK14 markers (hCV3001824, hCV2455926, hCV2455924) was mostly caused by an increase in the frequency of homozygous state (by about 10%) while the heterozygous frequency was reduced compared to the control sample. These three markers had very similar frequencies and were in high LD (Table 3.15), possibly representing the same haplotype. Only the

frequency of the marker hCV2455922 was reduced for the homozygous wild type state by about 10%, possibly representing a different haplotype.

Table 3.14: Allele and genotype association for severe disease phenotypes.

SNP marker	fistulizing disease association				bowel resection association				stenosis formation association			
	Allele		Genotype		Allele		Genotype		Allele		Genotype	
MAPK14	c^2	p-value	c^2	p-value	c^2	p-value	c^2	p-value	c^2	p-value	c^2	p-value
hCV3001824	10.98	0.00092	12.00	0.00248	5.64	0.01759	6.10	0.04743	5.42	0.01998	5.86	0.05344
hCV8852251	1.35	0.24876	1.51	0.47146	1.04	0.30668	0.97	0.61576	0.03	0.87652	2.63	0.26916
hCV2455932	1.64	0.20432	1.82	0.40314	3.13	0.07721	3.50	0.17375	3.24	0.07237	4.31	0.11580
hCV2455926	12.86	0.00034	13.50	0.00118	6.97	0.00835	7.37	0.02510	7.25	0.00715	7.39	0.02484
hCV2455924	11.42	0.00073	12.19	0.00226	5.38	0.02045	5.89	0.05243	6.51	0.01077	6.76	0.03410
hCV2455922	9.52	0.00206	9.52	0.00857	6.67	0.00983	6.76	0.03404	9.47	0.00212	9.42	0.00899
hCV2455919	4.43	0.03545	4.26	0.11887	3.22	0.07292	4.08	0.13009	1.42	0.23713	1.71	0.42684
MAPK13												
rs763096	5.02	0.02498	5.08	0.07855	6.59	0.01029	6.61	0.03670	4.11	0.04262	4.06	0.13111
rs2007683	0.02	0.89209	0.02	0.98869	0.24	0.63718	0.31	0.85718	0.09	0.78048	0.12	0.94135
rs2815812	3.83	0.05029	4.99	0.08239	3.97	0.04625	4.69	0.09572	4.21	0.04028	4.22	0.12109
rs2071863	10.63	0.00112	13.69	0.00107	8.98	0.00276	9.96	0.00687	12.98	0.00032	13.91	0.00095
TREM1												
rs2234246	0.01	0.93773	1.52	0.46852	2.03	0.15788	1.92	0.38395	2.30	0.13119	2.53	0.28292
rs1817538	2.01	0.16008	4.20	0.12216	8.47	0.00367	8.06	0.01775	5.14	0.02335	5.59	0.06104
rs1351835	0.80	0.37758	3.51	0.17309	5.62	0.01779	5.54	0.06252	4.72	0.02982	5.59	0.06110
rs1817537	0.90	0.34479	3.14	0.20802	4.88	0.02722	4.80	0.09063	4.63	0.03135	5.64	0.05949
rs2234237	0.00	0.96161	0.00	1.00000	0.11	0.75234	0.19	0.90705	0.15	0.69704	0.31	0.85752
prom-G217A	1.16	0.28280	1.69	0.43041	1.41	0.23928	1.99	0.36972	0.18	0.67686	0.87	0.64650

In the analysis of association with stenosis forming disease 303 affected individuals were included. The association was reduced compared to the full population. Only one marker from MAPK13 (rs2071863) showed a similar high association, than in the full population with a slight increase for the genotypic association (Table 3.14). The frequency in the affected sample population was increased by over 10% towards the frequent homozygous state. For the disease severity category of partial resection of the bowel 263 cases were included in the case control analysis. Similar to the analysis of the formation of stenosis, the association in the case control analysis in this category was weaker compared to the IBD or CD cohort probably due to the reduced sample size (Table 3.14). The TDT association to all three categories of disease severity did not reveal major differences to the association to IBD, CD or UC in the Transmit single marker analysis. One SNP marker from MAPK13 (rs2815812) had a higher significant association for all three severity categories (fistulae: 4.15, p-value: 0.0415, stenosis: 6.50, p-value: 0.0107, resection: 5.00, p-value: 0.0253). In the Genehunter single marker analysis no significant association was observed for all severity categories, even

though the rs2815812 reached a close to significant level (3.34, p-value: 0.067522 for resection).

Lewontin's D' values as a measure of inter-marker LD were high within the SNPs of each candidate gene indicating strong or even complete linkage of the markers. High inter-marker LD was as well observed between markers from MAPK14 an MAPK13 indicating their physical and genetic proximity. (Table 3.15 and 3.16). However some SNPs showed low D' values indicating that different levels of linkage between markers of the two genes was possible. This could be an indicator for different haplotypes represented by the SNPs or otherwise for differences in the ages of the SNPs.

Table 3.15: Inter-marker LD in MAPK14 and MAPK13.

hCV3001824	MAPK14										
hCV8852251	1										
hCV2455932	1	1									
hCV2455926	1	1	1								
hCV2455924	0.98	1	1	1							
hCV2455922	0.99	0.95	0.91	0.97	0.99						
hCV2455919	0.99	0.98	1	0.99	0.98	1		MAPK13			
rs763096	0.11	0.33	0.95	0.12	0.11	0.47	0.18				
rs2007683	0.39	0.06	0.81	0.38	0.37	0.57	0.02	0.99			
rs2815812	0.72	0.96	0.86	0.72	0.70	0.70	0.52	1	1		
rs2071863	0.93	1	0.87	0.96	0.93	0.98	0.41	0.99	1	1	
	hCV3001824	hCV8852251	hCV2455932	hCV2455926	hCV2455924	hCV2455922	hCV2455919	rs763096	rs2007683	rs2815812	rs2071863

Table 3.16: Inter-marker LD in TREM1.

rs2234246						
rs1817538	1					
rs1351835	1	1				
rs1817537	1	1	1			
rs2234237	0.98	0.98	0.98	0.98		
prom-G217A	1	1	1	1	1	
	rs2234246	rs1817538	rs1351835	rs1817537	rs2234237	prom-G217A

3.3. Analysis of linkage disequilibrium structure

With the intention to reveal structures of linkage between genes and SNP marker that could obscure association to disease or distort the true location of association an initial region of 10 Mb that was part of the association map and covered the central MHC region was analysed for LD in 920 SNP markers in a small US American White population. Subsequently a 3.53 Mb part of that region was analysed in more detail in 5 populations, three from Europe (Germany, Norway, UK) the original population of US American whites and a US African American population.

In order to quantify the pair-wise genetic differences between the populations analysed, F-statistics of population subdivision (F_{ST}) were calculated using all SNPs available for analysis with the Arlequin program. An F_{ST} value of 1 signifies a complete isolation of the populations analysed while a value of 0 denotes a complete admixture. Except for comparisons involving the English population, all differences were significant at the 95% level (F_{ST}-values between 0.00392 - 0.07180). The English population was found to be significantly different only from the US African-American population, to all other populations with European origin the difference was not significant.

After calculation of pairwise D' values for the genotypes of 920 SNPs in the US American White population with 45 individuals a cluster analysis of LD was performed using an unweighted pair group method employing the arithmetic mean (154). All markers employed for this analysis had a frequency of greater than 5%. Fig. 3.7 depicts UPGMA clustering from the 10 Mb region covering the MHC on human chromosome 6p (NCBI sequence coordinates 25.3 - 36.3 Mb).

On the vertical axis the markers are not in the physical correct order but rearranged according to the clustering. It was tried to be as close as possible to the physical order through the turning of the branches in the dendrogram, but in the clustering process sometimes markers were clustered that were not physically neighbouring. The vertical axis details the cluster dendrogram with a scale of D’ levels shown at the bottom. The scale indicate the arithmetic mean of all D' values of markers within that branch of the cluster tree. Levels of pairwise D’ are colour-coded from dark grey indicating high LD to light grey indicating low LD (in the top left corner).

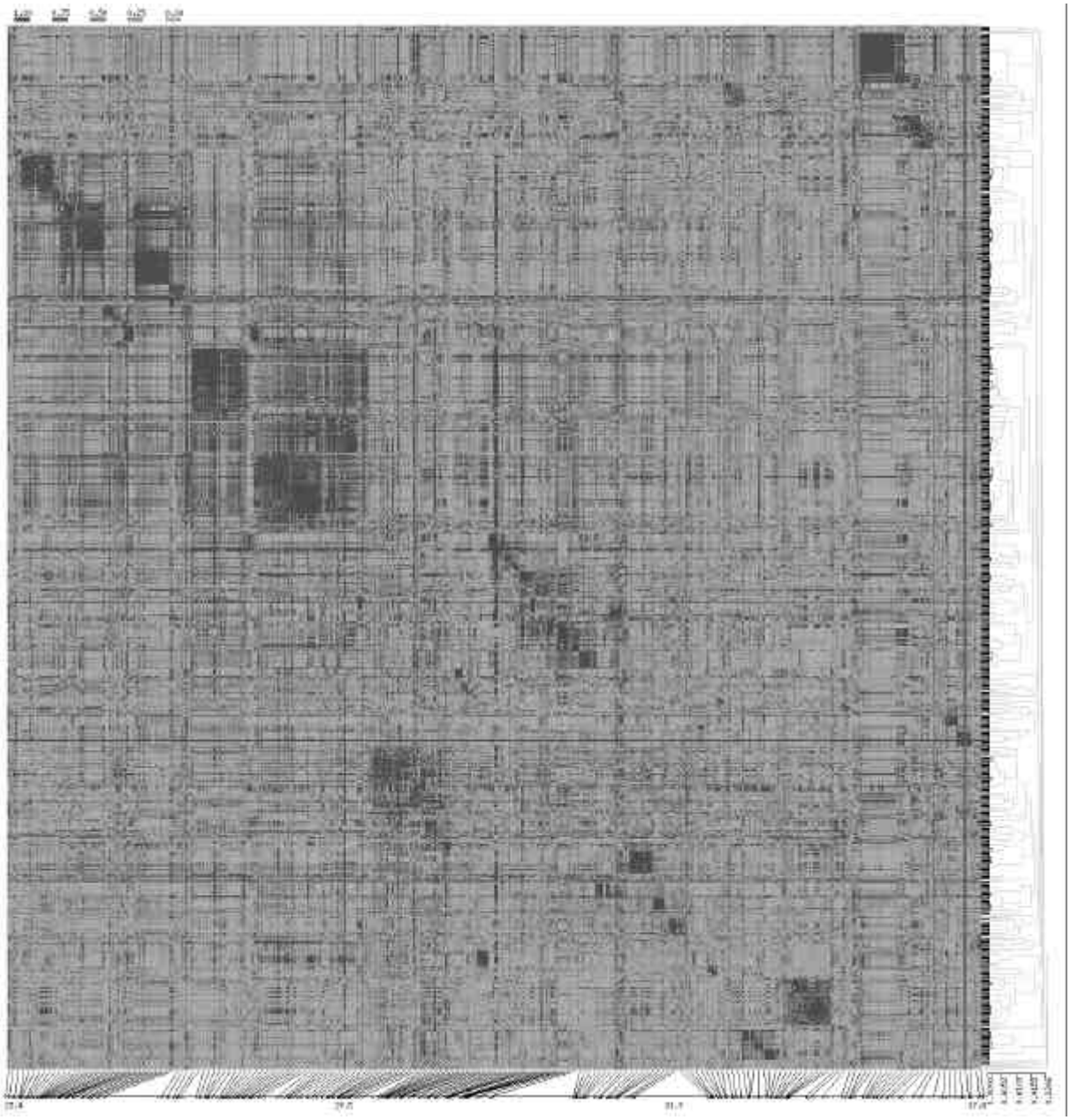

Fig. 3.7: Cluster heat-map of pairwise D' on human chromosome 6p21.A total of 920 SNPs (all with allele frequency of 5% minimum) located within 10 Mb (from D6S461 to D6S291) were genotyped in 45 white US American individuals. The graph relates the physical map of the region (horizontal axis, with every 5th SNP marked) to the order derived from UPGMA clustering of D' values (vertical axis). The vertical axis details the cluster dendogram with a scale of D' levels shown at the bottom. Levels of pairwise D' are colour-coded from dark grey to light grey as indicated (top left corner).

The degree of linkage structuring in this region was found to vary greatly, from highly organized parts with blocks of strong internal pairwise LD to sections of very low LD. Large blocks with high LD of more than 100 kb were apparent at positions 26 Mb, 27.5 Mb, 28 Mb, 32 Mb and 35 Mb; blocks of smaller size were found in the middle and in the most centromeric part of the region (e. g. positions 31.2-31.4 Mb and 36 Mb), but as well areas with high LD that can not be described as clear blocks and markers exhibiting high LD over

long distances could be observed. Generally the LD observed was higher in the telomeric half but no fixed size of blocks with high LD was observed. What was regarded as a block depends on the level of D' chosen as cut-off. In the cluster heat map presented here it was observed that at some instances there is a steep decrease of D' between one marker pair to the next marker pair, as seen in the block at 27 Mb or 36 Mb where the D' value decreased from 1 to below 0.25 within two SNPs. In other regions the change from high LD to low LD was in smaller steps as can be seen in the large high LD region at 28 Mb. This was reflected as well in the cluster dendrogram. The thick lines of the dendrogram show the D' clusters with a mean D' value of over 0.95. From the 10 Mb region described here a smaller region was selected for a more detailed analysis with more even distribution of the markers that was analysed in a larger population. The region chosen from the 10 Mb the contained all features described for the larger region, if so in a smaller scale. The MHC region is of special interest in respect to autoimmune and chronic inflammatory diseases and the 3.53 Mb from NCBI sequence coordinates 29.4 - 32.9 Mb showed to be representative in the terms described. The original markers where supplemented with 37 additional SNPs to minimize areas with low marker density and the total of 320 markers where genotyped in a 550 individuals large German population in order to increase the resolution of the analysis. Of the 320 SNPs located in this region, some 278 had a minor allele frequency above 5% in 550 German individuals.

For these markers, large and clearly apparent blocks with high LD were observed in the telomeric subregion (Fig. 3.8). The centromeric subregion showed areas of high pairwise LD, some regions of less well defined blocks, and areas with no clear LD pattern. Some marker pairs were found to exhibit high LD over long physical distances. These patterns were as well observed in the 10 Mb region. General size of the blocks with high internal LD was smaller in this region but different kinds of blocks could be observed as well: steep decreases of LD from one marker to the next, a slow reduction of LD from a centre of high D' and blocks of high D' values that were interrupted by makers with low D' values. The differences in the appearance were mainly caused by an alternative turning of the branches, which reflects the physical order of the SNPs better but not exactly. This could be due to the higher number of individuals. The clustering was similar to the initial analysis with the difference of the additional markers. The genes that are included in this region can as well be seen in the Fig.3.10a, which shows exactly the same region. An obvious correlation of the classical HLA genes with a region of high or low D' values was not observed.

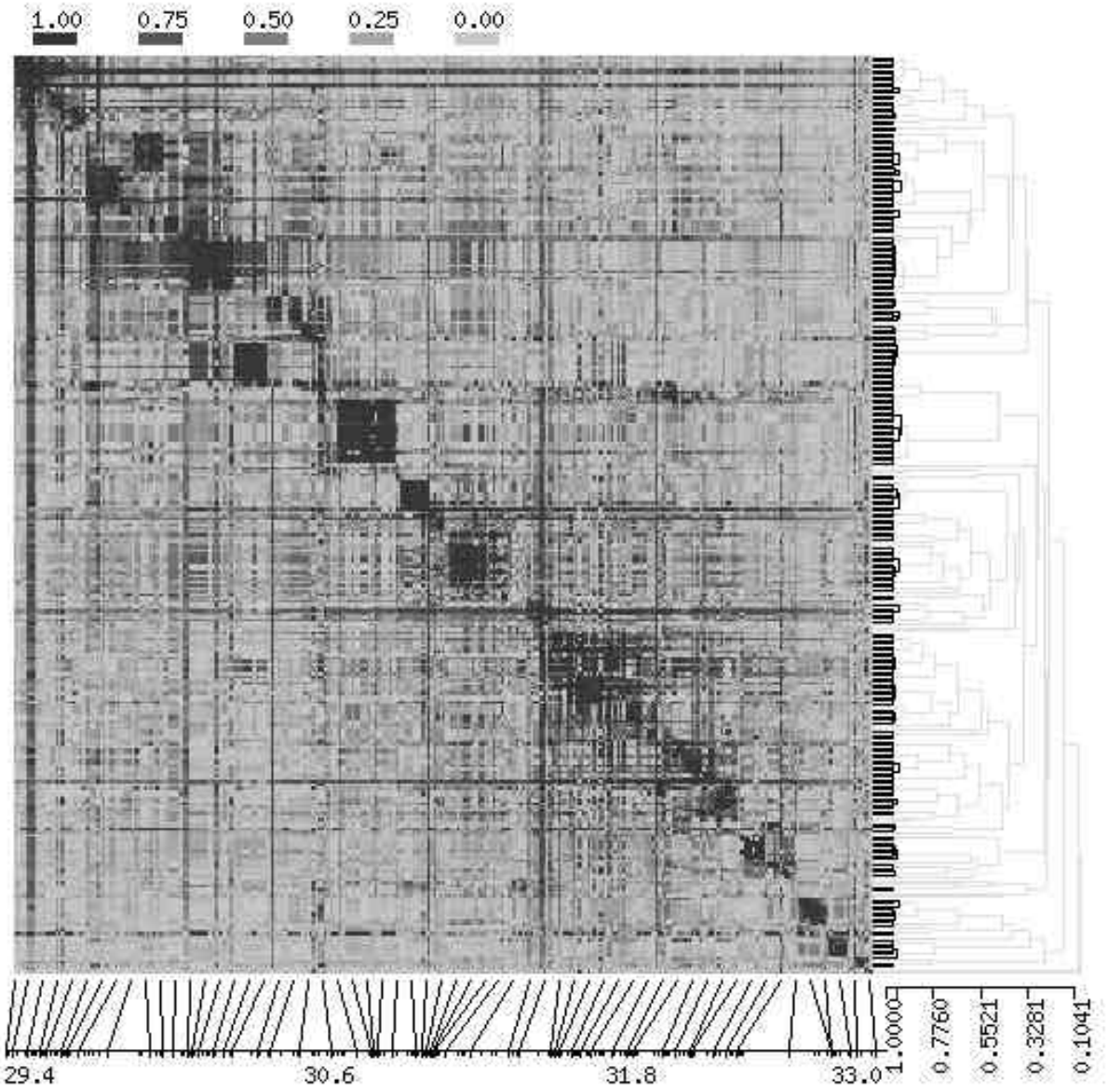

Fig. 3.8: Cluster heat-map of pairwise D' in a representative 3.53 Mb region including the human MHC. A total of 320 SNP markers were genotyped in 550 German individuals. 278 SNPs with minor allele frequencies >5% were analysed. For details: see legend to Fig. 3.7.

Cluster analysis adopting a 20% allele frequency cut-off (199 markers) resulted in a substantial reduction of long distance LD, but gave a similar quality and distribution of haplotype blocks (Fig. 3.9). The major characteristics of haplotype structure were found to be similar in all population samples (Figs 8.4 - 8.6, appendix 8.1.4). The degree of haplotype structure was not related to marker density (Fig.3.7, Fig. 3.10a).

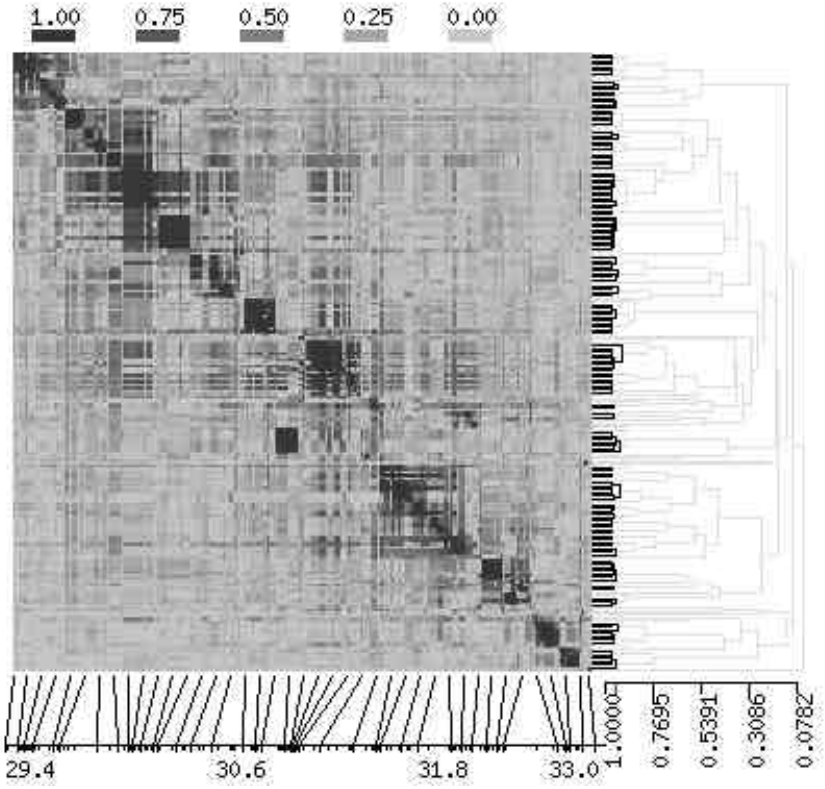

Fig. 3.9: 20% cut-off cluster heat-map of pairwise D' in the 3.53 Mb region. 199 marker had a frequency of above 20%. For details: see legend to Fig. 3.7.

An analysis of the density of the markers was performed by plotting the amount of markers within a 500 kb sliding window along the physical map (Fig. 3.10a). This 500 kb sliding window was used in the correlation analysis again and by this way the results became directly comparable. On top of the figure a selection of genes is placed according to their position on the physical map. Pearson correlation coefficients between D' values in the sliding 500 kb window in 45 randomly drawn individuals from the different populations were calculated using the German individuals as reference population. Within the German population a random sample of 45 individuals was drawn 1000 times and the correlation coefficient was each time calculated with the remaining 505 individuals as reference. The average of the 1000 was taken and the 95% confidence interval plotted. The analysis reflected the results from the cluster heat-map showing a high correlation where the high LD block structure was observed in the cluster analysis of the whole German sample (Fig. 3.10b).

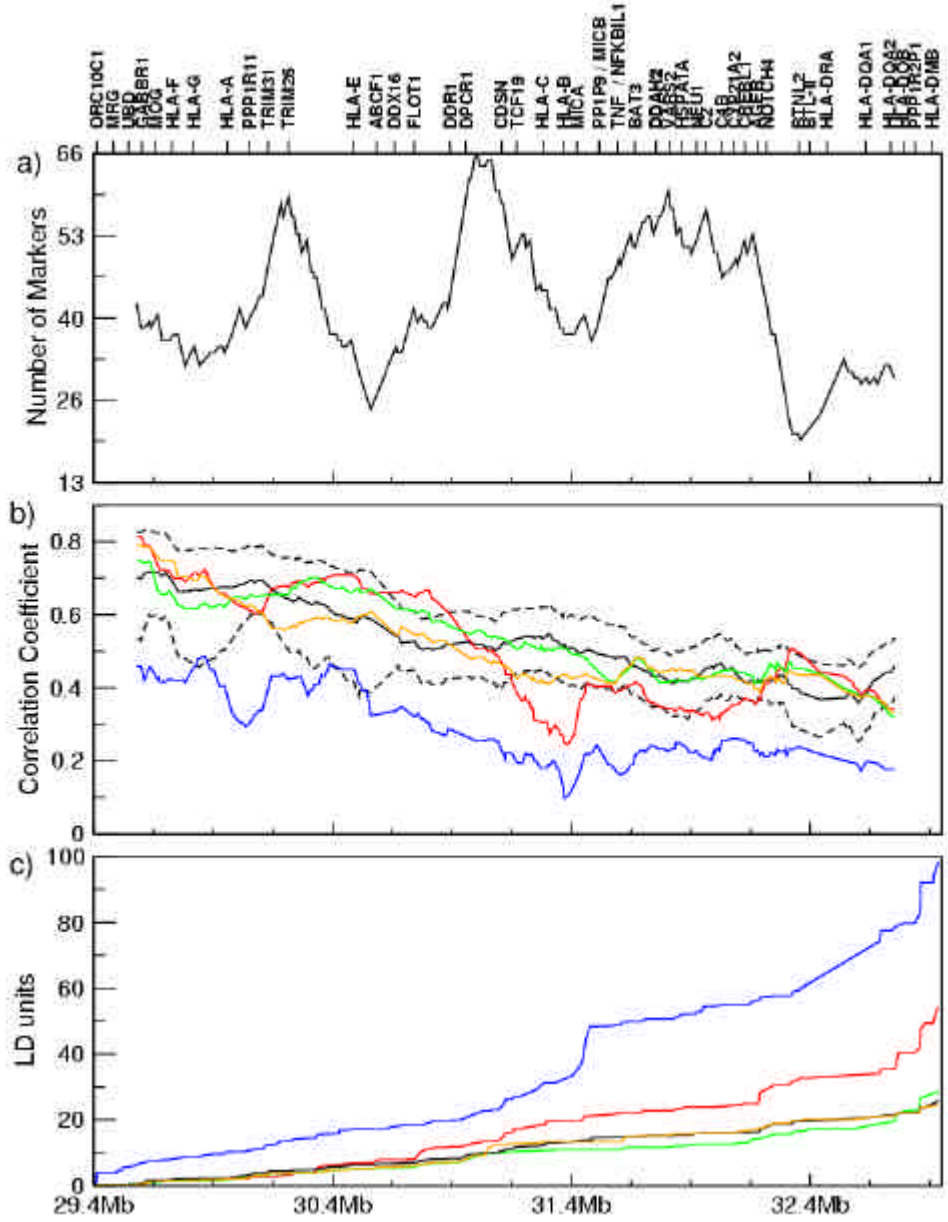

Fig. 3.10: Analysis of LD in a 3.53 Mb region covering the human MHC (NCBI sequence coordinates 29.4 - 32.9 Mb).
Panel A: Density and map of SNPs genotyped. The total number of SNPs in a 500 kb sliding window (vertical axis) was plotted against the location of the centre of the window on the physical map (horizontal axis). **Panel B: Correlation of D' between populations.** D' values were calculated between pairs of SNPs with allele frequency above 5% in 45 randomly drawn individuals from each population (UK White, 270 SNPs – green, Norwegian White, 267 SNPs – yellow, US American White, 221 SNPs – red, US African-American, 209 SNPs – blue) and compared to D' values obtained in the German White population (278 SNPs) in a 500 kb sliding window (vertical axis) 95% confidence intervals (dotted black lines). **Panel C: Analysis of LD units.** (by Francisco De La Vega at Applied Biosystems, Foster City, California 94404 / USA) Metric LD units (vertical axis) are obtained from a correlation between the distance of SNPs on a LD map and distances on a physical map (horizontal axis). Regions of strong LD result in sections of the curve where LD increases slower than physical distance (i.e. the curve flattens), while regions of high recombination activity are characterized by a disproportional increase of LD units over physical distances. Identification of populations by colour-coding is described in Panel B.

A high degree of correlation, indicative of larger high-linkage blocks, was consistently seen in the telomeric subregion (29.4 - 30.6 Mb) whereas correlation coefficients declined toward the less structured, centromeric end (31.2 - 32.2 Mb). Correlation coefficients between the German and the other populations followed a similar trend: with few exceptions, all values from the Europeans and the white US sample fell within the 95% confidence interval defined in the German sample. These findings suggested a strong conservation of LD structure between populations. This notwithstanding, the coefficients from the US African-American sample fell outside the 95% confidence interval although a decrease was also seen to occur from telomere to centromere. This indicates the higher similarity between populations of European origin reflecting the common history. An influence of the marker density on the correlation could not be observed.

Metric LD units

This analysis was performed through Francisco De La Vega at Applied Biosystems, Foster City, California 94404 / USA. due to reasons of consistency it is not separated from the study.

Linkage between markers is restricted through recombination. Therefore the analysis of recombination between markers over the region was also assessed in the form of metric LD units (LDU) (155) with the LDMAP program (Fig. 3.10c). This measure is based on the decay of LD calculated with the aid of the “swept-radius” which has been suggested as the maximum (physical or genetic) distance over which useful LD can be expected (155). In line with the cluster- and correlation-based analyses of D’, LD was found to be high in the interval 29.4-30.78 Mb, resulting in a plateau in the LDU plots of all populations (Fig. 3.10c). In the centromeric subregion, a higher rate of recombination became evident in the form of steeper increments. Consequently, the number of LDUs per physical distance unit was higher in the 3’ than in the 5’ half of the plot. After positions 29.4-30.78 Mb, the LDU plot of the white US American sample departed from the European samples, and differed particularly towards the centromeric end of the region. The total LDUs attained in the 3’ half of the region by the white US American sample was approximately 70% higher than for the European samples. The US African-American sample exhibited a substantially different LDU pattern in that it attained a 3.5 times higher total LDU for the whole region than the European samples, and an 80% higher total LDU than the white US American sample. The overall pattern of steps and plateaus was the same in all samples, although the US African-Americans exhibited a prominent increase in cumulative LDU at distinct positions, particularly at 31.48 Mb.

4. Discussion

The aim of this study was to further confine a previously identified linkage region for inflammatory bowel disease and the disease categories Cohn's disease and ulcerative colitis and from this to identify functional and positional candidate genes with the final aim to eventually identify one of the genes carrying a causative mutation for IBD. Due to the results from the association map and the experience with the genes analysed as positional and functional candidate genes it became necessary to refocus the work with the aim to reveal underlying structure of linkage disequilibrium between markers in the region of interest that could distort the true association locus or probably account false positive association results

4.1. Association mapping

Association mapping employing diallelic SNP markers was performed in the previously identified linkage region on chromosome 6p (bordering microsatellite markers: D6S461 to D6S271). Therefore a 22.8 Mb large region including the MHC was covered with 142 diallelic SNP markers and genotyped in a large population sample of monoplex and multiplex families and unrelated controls. The SNP markers were tested for allele and genotype association with the disease phenotypes IBD, CD and UC in the case control study design and for TDT association in the single point and multipoint analysis for up to 3 markers. Several association leads were observed with each of the methods, however not entirely overlapping. Markers with the highest associations in the leads were mostly the same in the allele and genotype association in the case control study design, differences were observed for the disease specification for which the highest association was seen. While three major association leads in the regions of HSP70-HOM to HLA-G, SLC26A8 to MAPK13 and around TREM1 were identified in the case control association, the TDT analysis with Transmit had only one major association lead from HSP70-HOM to rs1035798 (AGER) and 4 minor peaks. All peaks were observed as well in the Genehunter TDT analysis, pronounced to a different degree. One peak was only present in the UC population. The others were relevant for IBD with a greater weight for CD phenotype.

While the overlap of some association results gives a good indication toward a potential location of a disease gene, the differences between the results achieved with the different methods of analysis need further enquiry. The use of diallelic SNPs as markers for association

fine mapping has been established by now as an adequate tool (158-160). Through the LD between physically close markers association to a disease causing mutation could be identified through an SNP that is located close if positioned on the same haplotype. Whether susceptibility genes can be detected is dependent on the degree of linkage disequilibrium (160). The presence of haplotype structures through LD between markers could as well give rise to association at a certain distance from the true stimulus. Therefore a thorough knowledge of the LD structure in the region of interest is needed.

Problems intrinsic to the use of SNPs as markers for association mapping are the allele and genotype frequency of each marker (160), which could vary for ethnically different populations, and as well where undetected population stratification is present. If there is a large difference in allele frequency between the disease variant and the marker, the disease variant cannot be detected due to the different extent of LD and the difference in the haplotype (160). Additionally markers with a low allele frequency are of low information content as most of the genotypes analysed are identical be state (ibs) and not necessarily by descent (ibd). In the TDT analysis it has to be taken into account that only heterozygous parents are informative as only in this case it can be distinguished which allele was transmitted. In the situation of an SNP at HWE this means that at least 50% of the parental samples are not included in the analysis, which increases for rare alleles. This reduces the effective population size in the TDT analysis, which could either mean a substantial loss of power to detect association or the overestimation of association on the basis of small numbers of alleles actually transmitted or not transmitted. This effect might be a reason for the differences between the case control and the TDT association observed in this study. The phenomenon was more pronounces in the multimarker analysis where due to the larger number of possible haplotypes the frequency of some haplotypes is very low, even though the major two-marker haplotypes associated to one of the disease categories were not affected by this problematic. A similar situation is possible for the direct typing of classical HLA genes where a great number of alleles reduces the effective sample size for each allele resulting in overestimation of positive associations.

The distribution of the SNP marker in the association map was not uniform. While a central region from the chromosomal position 29.4 Mb to 37.5 Mb has a marker density of 80 kb and a median distance between markers of 36 kb, the density outside this region was lower. This inconsistency and the variability in the distance between two markers were partly caused by

the selection criteria. Several SNPs were not confirmed during the sequencing verification, especially SNPs from dbSNP were to a high degree not detectable in the sequencing population used for verification. Additionally some assays were withdrawn from the analysis because of bad allelic discrimination, a high rate of unexpected null-alleles or a great number of Mendelian inheritance errors. The variability in the distance of the SNP markers entails problems with the association mapping. Association that peaks in the uncovered part of the map might not be detected and the centre of a detected association might not be defined precisely. The fidelity of the typing technology employed itself was high as was seen in a marker that was genotyped twice in the complete association cohort. The same alleles were assorted to the individuals involved in both tests.

Crucial for the success of a mapping experiment is the correct positioning and distance of the markers in the genetic map. The region investigated had been subject to intensive sequencing (3). Still the exact positioning of the SNP markers and genes changes with every new release or between different sequence assemblies. The problem is more pronounced regarding the microsatellite markers employed for the initial screening of the chromosome 6. The different map constructs (e.g. Genethon, Marshfield, NCBI_RH etc) place the markers at different positions and distances to each other. For the selection of the region from which the markers were selected the NCBI sequencing positions were employed and SNPs between these were included into the analysis. As recombination within the SNP marker map was not significantly higher than expected from the physical distance, it can be assumed that the positioning of the SNPs to each other and the distance between the markers employed for the analysis did not alter considerably from their true location.

An analysis of the linkage with disease in the SNP markers was performed and reached NPL score up to 3.47 in the IBD. Even though this reaches not the suggested standard of 3.6 (41) the value indicated strong linkage to IBD, taking into account, that the diallelic SNP markers are less informative than microsatellites, and the comparably small sample size of ASPs. The NPL score decreases only toward the centromeric region markedly below 3 while to the telomeric region it stays high to the end of the map. This might indicate that a greater linkage might be undetected telomeric of the region investigated.

Three regions with increased association were identified through the case control analysis and were partly confirmed in the TDT association. These association leads were used as indicator for an area where a disease relevant gene could be located.

4.2. Candidate genes

HLA-DPA1 is one of the genes that could be a candidate gene due to the association in the surrounding markers and the antigen presenting function as a HLA gene. Few studies have investigated HLA-DP. Most association studies involving the classical HLA genes focused on HLA-DR, or DQ. Some of these studies included the HLA-DPB1 and did not report significant results (103, 104, 113). In the one study analysing the DPA1 gene, a positive association was seen between CD and DPA1*0201 (114) in the Japanese population. A large European population of multiplex families with IBD affected children was sequence analysed at three polymorphic sites in exon 2 of the HLA class II gene DPA1. Results were compared to a South African family population and a South Korean case control population. Two of the polymorphisms are causing amino acid exchanges in the N-terminal domain of the HLA-DP antigen. The three loci form six shared epitopes according to the connected HLA-DPA1 alleles (Table 3.6). No other polymorphism on the exon 2 was considered as these three loci represent the most common nucleotide exchanges (22). The positive association described for the HLA-DPA1*0201 allele in Japanese CD patients (114) was not replicated in this study. Although this might be due to the ethnic distance of the investigated population samples to the Japanese population (22) even in the Korean study population, which is similar in the HLA-haplotype distribution, no similar result was observed.

Suggestive but weak evidence for association between the shared epitope representing the HLA-DPA1*02021 / *02016 alleles and CD was observed in the German family cohort. This was not seen in the smaller patient populations from South Korea or South Africa. This finding may point to a conserved haplotype existing in the German population that includes the allele DPA1*02021. A similar situation probably is present in the Japanese population explaining the positive results for the DPA1*0201 (114). In the South African population a marginally significant association for the non-coding polymorphism to ulcerative colitis was seen for the shared epitopes on exon 2. As this was not seen in any of the other populations, the mixed ethnic background and the comparatively small size of our South African ulcerative colitis study population could account for this result.

The incidence of CD and UC in the South African population, especially among persons of coloured ethnic background is low (161). In the Korean population, the situation is similar (95). Recruitment of affected individuals in these populations therefore was focused on unrelated patients to serve as a control for positive findings in the German population. The recruitment of multiplex families, which would be appropriate for a comparison to IBD affected families of the European population sample, was not feasible.

The large size of the European study population and the possibility to compare to populations of highly different ethnicity made it possible to identify gene variations and shared epitopes even at a low impact level. The mainly negative findings are a strong indicator that the HLA-DPA1 gene is not directly involved in the etiology of IBD. In order to exclude the HLA-DPA1 from the candidate genes, a complete analysis of the DP antigen including the HLA-DPB1 would be necessary, as common haplotypes are known (22) and the functional interaction as antigen is only given through the α-β heterodimer on the cell surface. While population size in most HLA-typing studies is a problem due to the large number of possible variants scaling down the effective sample size, this was not critical in this analysis. However the reduction to 3 polymorphic sites is not representative for the complexity of the MHC either. For further studies of classic HLA genes it might be necessary to include all known alleles in each of the HLA class II genes and the use of more clearly defined disease subgroups (113).

Besides the sequence genotyping of the HLA-DPA1 4 other genes were of interest as positional and functional candidate genes. The three genes MAPK14, MAK13 and BRPF3 were located within one of the association leads of the case-control association study. In the TDT analysis the association did not reach the significance level. The situation was similar for the TREM1, which was situated within a smaller association lead centromeric.

MAPK14 is involved in a range of inflammatory processes. It has an important part in stress related transcription and cell cycle regulation Activation of the protein is triggered through environmental stresses and proinflammatory cytokines. While relevance of p38-α for inflammatory bowel disease has been shown, especially with regard to the TNF-α related pathophysiology (162, 163), the role of the kinase in the disease mechanism is yet not understood in detail. A significant rise in p38-α activity, but not expression levels, could be

observed (162) which would indicate an altered binding or phosphorylation site in the molecule itself. An alteration of phosphorylation sites or a modification of binding sites for reaction partners would require coding mutations in exons.

In this study the exons and the potential promotor region of the gene locus of MAPK14 were sequenced in 60 chromosomes, some exons were reanalysed in another 94 chromosomes. No exon polymorphism was found and the polymorphisms described in the dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) as exonic could not be confirmed in the individuals analysed. The alternate spliced exon 9 and 10 used in MAPK14 and CSPB1 have the same length of 81 basepairs and about 30% identity. The position of the alternate exon use is marked through a double sequence in cDNA analysis that could falsely be taken as SNPs (139). The absence of SNPs in the exons eliminates the possibility of an amino acid change in one of the active sites. The findings confirm that MAPK14 and its variants in human and other vertebrates are highly conserved (164). Cross species comparisons between chimpanzee and human revealed only three base exchanges in the translated region (139). This finding supports the pivotal importance of the enzyme in the cellular signalling pathway. The three known splice variants of the gene and the absence of SNPs in the coding sequence suggest that alteration of the gene through conventional mutational processes would have been deleterious in the evolution. Alternatively the use of splice forms could have evolved.

The analysis of intronic and 3'UTR SNPs revealed association with IBD in the case control analysis. The support for this was very weak and not significant for the TDT analysis. Neither the stratification for nor against carriers of the three CARD15 mutations (3020insC, R702W, L1007P) (63-65) did alter the association results except for a reduction that could be expected due to reduced sample size and power. In the analysis of severity factors for IBD development of fistulae, the formation of stenosis and a previous resection of parts of the bowel as described through the affected individuals several markers from MAPK14 showed slightly increased association with the fistulizing disease despite the reduced population sample. This could be due to a more stringent phenotype but the effect was not observed in the TDT analysis. None of these SNPs could be identified as a classical splice site or splice enhancer. Otherwise the functionality of intronic SNPs is very limited and difficult to prove. The differences observed in the activity of the MAPK14 could be explained by a change in efficiency of an upstream signalling pathway component of MAPK14.

The MAPK13 kinase is closely related to MAPK14 with a similar function and activation and the results observed were similar. No exonic SNPs were identified in 94 chromosomes. The gene was located within the same association lead as the MAPK14 observed in the case control study design, and had no significant association in the TDT analysis. While the stratification results for NOD2 carrier showed no significant association, one marker of the MAPK13 (rs2071863) seemed to support the complete association observed in the full cohort. This result was still significant after the permutation analysis. The marker exhibited a shift toward the frequent homozygous genotype of about 10% reduction of the heterozygous state. Matching to this is the observation that most of the association of that marker was observed in the CD population as the NOD2 mutations account for part of the affection with CD. Still, in the TDT analysis the association was not significant. The same marker was observed with the highest associations in all three analyses for severity phenotypes, but the association was not carried fully. Again no direct function could be assigned to this SNP.

For the TREM1 gene which is expressed in monocytes and neutrophil cells (145), and is involved in monocyte/macrophage- and neutrophil-mediated inflammatory responses. The sequence analysis identified three at that time novel SNPs, two of which are by no described as rs2234246 and rs2234237. While rs2234246 is in the 3'UTR only rs2234237 is an exon polymorphism with an amino acid exchange from T to S, this was the only marker not exhibiting any significant association in IBD, CD or UC. The third SNP detected through sequence analysis is in the potential promotor region. Association for all other markers was slightly stronger in the UC population regarding the case control design. In the TDT analysis none of the markers reached significant association. The case control association was completely carried by the NOD2 negative individuals, which agrees with the stronger association in UC, while none of the severity factors were observed to be of major association.

One possible cause for the association signals would be linkage disequilibrium with a causative mutation with which the SNPs with positive associations share the same haplotype. Markers of both genes MAPK14 and MAPK13 were tightly linked within the gene but as well between the genes. Therefore they could represent the same haplotype associated with disease. Through the different allele frequencies the association might be altered to a certain degree. A less frequent allele usually indicates a younger SNP that displays LD in a higher degree to the surrounding marker and over a greater distance. An allele with the same

frequency as the disease causing mutation will represent the association more precisely. One gene that is close to the association lead seen in the SNPs of the MAPK14 and MAPK13 was the BRPF3. The positive association in the SNPs analysed could have been caused by LD with a potential causative mutation in BRPF3. However the sequence analysis of the BRPF3 gene did not reveal any SNPs in the cDNA in the individuals analysed.

Other possible explanations for the positive associations observed could be a variety of mechanisms. Even though the analysis of severity factors did not increase the association signals, the definition of the phenotype analysed surely is of critical importance. Several measures had been taken to make the disease definition of IBD as well as UC and CD consistent. Nonetheless, the spectrum of phenotypes in these definitions is large regarding the intensity of the diseases, the location of inflammation, the bandwidth of characteristics and the extra-intestinal manifestations (60, 61, 119). Especially in regard to the inflammatory diseases associated or linked to the MHC region which exhibit a considerable degree of overlap of certain manifestations like the inflammation of skin, joints or lungs (57), overlapping with the descriptions of psoriasis, rheumatic arthritis and sarcoidosis, there is a need to review the strategies adopted to identify causative mutations. Many of these diseases have more than one potential susceptibility gene. Certain characteristics present in different diseases could possibly be caused through the same gene in each of the diseases. In IBD about 18 % of the genetic risk to develop CD could be attributed to the C-insertion mutation in the NOD2 gene on chromosome 16 (63-65). Further analysis revealed that the NOD2 mutations determine ileal disease only (165, 166). For the narrow disease description peripheral arthropathy of IBD it was demonstrated that a precise definition of phenotype could lead to a highly significant association with a gene (167). The collection of a cohort that complies with these conditions would need an in-depth definition of disease phenotypes and an intense screening of affected individuals, to achieve a population size that gives enough power for the analysis. Besides the other difficulties, the penetrance of the disease variant has to be considered (160). The intensity of penetration is difficult to estimate in a polygenic disease and no single model can be applied more significantly than others (91).

Other important factors in the description of disease are the environmental influences. In twin studies it was observed, that even in monozygotic twins the concordance rate was 6.3% for ulcerative colitis and 58.3% for Crohn's disease (80). The role of the environment has been discussed, and especially the influence of modern industrialized life-stile seems of great

importance (66, 81). The presence of environmental factors triggering the development and possibly the phenotype of the disease add another variable to the analysis.

Further evaluation in the candidate genes analysed herein would include the examination of splice variants and their distribution in the different disease categories compared to unaffected individuals and the probably the identification of additional splice variants. This would as well include mutation detection in the full sequence including the introns, which for example for MAPK14 stretch over several kb. Furthermore other genes within the association leads could be as well functional candidates and need further investigation.

4.3. Linkage disequilibrium structure

An organization of the human genome into conserved haplotype blocks, interrupted by historical recombination “hot-spots” or (less active) “warm-spots”, has been proposed repeatedly (25, 48, 56). At the same time, a number of methodological approaches to identify haplotype structure and to delineate recombination hot-spots have been expounded (51, 53, 56). A ubiquitous presence of conserved blocks would allow the construction of LD maps of the human genome, facilitating characterization of genetic diversity between and among populations. Since previous results on MHC diversity and its presumed evolutionary history appeared to be incompatible with a strict organization into haplotype blocks, this region was investigated in more detail in different European and US-American populations.

The construction of UPGMA trees based upon pair-wise D’ values and their comparison to physical marker maps allow the assessment of haplotype structure without any *a priori* hypothesis about the nature of such structure. The analysis of a 10 Mb region on chromosome 6p, including the MHC, demonstrated the coexistence of different LD patterns in closely linked genomic regions. Although recombination hot-spots and haplotype blocks emerged in all population samples analysed, a universal and uniform haplotype structure was neither observed in the 10 Mb region nor in a 3.53Mb peri-MHC region analysed in more detail. Furthermore, recombination events were not confined to “hot-spots” even though the latter could be inferred easily in areas of conserved haplotype structure. The distance, up to which useful LD was still detectable within haplotype blocks, if present, was approximately 100 kb and corresponded to the typical block size suggested before (55).

Correlation in D' between populations can be expected to be low in areas of high haplotype diversity. In this sense, the white European populations exhibited a qualitatively similar distribution of haplotype structure throughout the genomic region. Substantial differences were only observed in the comparison to the US-African American sample whereas the white US-Americans were relatively close to the Europeans for most of the analysed region. This finding is likely to reflect the known historical relationship between the populations and their migration histories.

Similarities between populations were generally greater in regions with high LD, indicating the presence of a limited number of haplotypes generated by a small number of historic recombination events. This supposition was confirmed by the analysis of LDU as the correlation in D' was found to parallel changes in LDU. Where historic recombination rates must have been high, correlation coefficients were low. The genomic position at which the LDU plot of the white US Americans diverted from the Europeans coincided with a sharp decrease in D' correlation coefficient, placing the Americans outside the 95% confidence interval of the German intra-population comparison. One explanation could be recent admixture of the white US-American population with other European or non-European populations that broke up "old" European haplotype blocks. An even more pronounced trend towards wide-spread and frequent historic recombination events was seen in the US-African American sample, a finding that corroborates previous studies (53, 168).

The density, allele frequency and age of SNPs all have an influence on the nature of the haplotype structure that can be detected. For the present analysis, therefore a 5% frequency cut-off was adopted (leading to the exclusion of 6% of markers) to reduce the number of evolutionary young markers that led to long-range LD. A high number of comparatively young SNPs with long distance LD and allele frequencies above 5% may be expected to exist in the MHC region due to its propensity to frequent recombination events and the selective pressure towards higher variability. A reanalysis adopting an allele frequency cut-off of 20% (32% of markers excluded) led to a further marked reduction in long distance LD, without changing the principal findings. For the density of the markers there could no direct dependence to the correlation or the LDU be observed. However even the 500 kb window with the lowest marker content had 18 markers. A lower density than this could still influence the level of LD detectable especially as with a decrease of density the risk of very large gaps rises.

Sample size also has an impact on the statistical ability to describe haplotype structure. The results suggest that a sample of 45 individuals would be sufficient to delineate major haplotype structure in areas of low recombination, but that larger samples and higher densities of SNPs might be required to resolve short haplotype structure in areas of high recombination activity. Therefore, every SNP available may have to be included in the genotyping and analysis strategy in those areas in which haplotype structure was not detected by the present study.

The physical and biological explanation of the local variation and the interregional differences in LD structure is unclear. A link between recombination rate and occurrence of mutations has been suggested by some studies (54, 169). This should be found, when high variability and a fast turnover of new variants are of biological advantage in a given genomic region as it has been suggested for the MHC region. However, the classical HLA genes characterized by high levels of variability were not linked to areas of particularly high or low LD in our study. A higher degree of sequence conservation is usually observed for the introns rather than the exons of these genes, and the evolutionary mechanism underlying this phenomenon probably reflects the need for repeated recombination (170, 171). Similar patterns of LD distribution were also found outside the MHC, in the 10 Mb region analysed here and probably also exist in other chromosomal regions (172). Primary DNA structures as well as DNA topology have been discussed as potential factors promoting recombination (52, 173-175).

The results on LD distribution may prove useful in the efficient assembly of SNP panels for future association mapping of chromosome 6p-linked complex diseases. In contrast to previous assumptions, substantial parts of the human genome cannot be parsed into uniform haplotype blocks and recombination hot-spots. Differences in recombination and LD pattern have to be taken into account for the study of variability between populations, even when block modelling of the more conserved regions is undertaken. The sample sizes and SNP density needed has to be adapted to the expected level of structuring by haplotype blocks. Thus, regions with well-defined block structure could be represented by fewer SNP markers or on the basis of smaller population samples than regions without such structure.

In regions without haplotype block structure a high number of SNPs may be needed to detect association. The presence of long-range LD may result in significant problems in the

positional localization of the primary disease associated genetic variant. Genome wide haplotype mapping efforts will therefore result in a critical resource for the future of genetic mapping in phenotype diversity studies.

4.4. Conclusion

Association mapping with diallelic SNP markers is a useful tool to refine a previously identified chromosomal linkage region. Increased association could be detected with this method at 3 locations in the case control study design and 4 association leads were observed in the TDT analysis in families. Association detected through these methods is calculated on the basis of different assumptions. While association within a case control study design is based on identity by state of the disease allele and the marker used for analysis in the group of independent cases, the analysis of family based association identifies the transmission of alleles from parents to offspring. Both methods have difficulties and advantages. While a case control population is easier to assemble, the analysis of healthy parents minimises the influence of artificial differences. Due to the different kinds of analysis the resulting intensity and location of association might as well differ but still map the same signal, as the methods are influenced to a different degree by LD in the region of interest, the allele and genotype frequency and environmental factors. Limitations of using SNP marker and association as measure to determine the location of a potential susceptibility gene are the marker density and the structure of LD in the region analysed, but as well the frequency of the disease causing mutation. A certain marker density is necessary to map a region for association. The necessary minimal density was probably only achieved in the internal more densely covered central region of the map used in this study, so that association on the margins might not be represented well enough. The signals in the central part of the map were clear and useful for further analysis and the fact that none of the SNPs analysed in the candidate genes is the causative mutation does not reduce the possibility to find a disease locus within the associations detected. Several other functionally interesting genes are located in the area. Important for the interpretation of the data from the association is the knowledge about underlying LD structures in the region analysed. Where LD is strong the association signal might be carried over a long distance and the representativeness of the association depend as well on the frequency of the marker. Taking these aspects into consideration, the association mapping could be a useful tool in further disease gene identification.

On the basis of the association leads some functionally interesting genes were analysed in more detail. As a classic HLA gene the HLA-DPA1 was analysed on three polymorphic sites in the second exon. The data presented in this study suggest, that the HLA-DPA1 locus itself is not a major genetic risk determinant for IBD in any of the three populations described here. The not significant population-specific association signals might point to population-specific conserved haplotypes that may carry true susceptibility mutations.

In four other genes from the association leads mutation detection on the exons and the potential promotor regions were performed. Only in the TREM1 gene a polymorphism was found in the second exon. None of the polymorphisms analysed in MAPK14. MAPK13 and TREM1 could be identified as disease causing. In the BRPF3 gene no polymorphism was discovered in the analysis of cDNA and no intron SNPs were known. The high conservation of the MAPK genes, which might be entailed through their relevance in the cell signalling and the presence of several splice variants might make it necessary to investigate other factors influencing the known increased activation of MAPK14. This could be additional splice variants with splice signals affected through intronic SNPs or binding sites for enhancer proteins. The observed association points to a variant causing disease in linkage disequilibrium with the SNPs analysed.

With the results from the association mapping experiment and the analysis of candidate genes it became obvious that the previously existing descriptions of haplotype structures in the MHC were not representative and do not sufficiently explain the observed results. Therefore the region was analysed for LD structures in 5 different populations. In the 10 Mb region and the internal 3.53 Mb region investigated a great variety of linkage disequilibrium structures were observed. High LD levels and clear haplotype block structures were found to intermingle with areas of low or medium LD, both inside and outside the MHC. The pattern and characteristics of LD were similar in all European populations and the US American white population. The African-Americans, despite exhibiting a similar distribution of LD over the region, are nevertheless characterized by a higher degree of haplotype diversity. These findings are likely to have a strong impact upon the future design of systematic association studies of chromosome 6p-linked human diseases. In regions without haplotype block structure a high number of SNPs may be needed to detect association. The presence of long-range LD may result in significant problems in the positional localization of the primary disease associated genetic variant. Therefore, even studies of functional candidate genes will

benefit from a map detailing the regional distribution of haplotype block structure. Genome wide haplotype mapping efforts will therefore result in a critical resource for the future of genetic mapping in phenotype diversity studies.

5. Summary: Development of an SNP map in the peri-MHC Region on the human chromosome 6 as a tool to identify candidate genes for inflammatory bowel disease

The human MHC region is located at the chromosome 6p21 and encodes for a great number of genes that express proteins with high relevance in the immune response and regulation. This includes the classical HLA genes, which are important in the recognition and presentation of antigens foreign to the body, but as well members of the cytokine signalling network, like tumor necrosis factor (TNF-α) or mitogen activated kinases (MAPK14). This makes the region an interesting target for the identification of susceptibility genes for autoimmune or chronic inflammatory diseases. Linkage to chromosome 6p21 and the MHC region has been established for several inflammatory and autoimmune diseases as for example for inflammatory bowel diseases (IBD). Inflammatory bowel disease, which can be divided in two phenotypes Crohn's disease (CD) and ulcerative colitis (UC), is a relapsing chronic inflammatory disorder of the gastrointestinal tract influenced by complex genetic, immunological and environmental factors.

The first aim of this study was to develop a single nucleotide polymorphism (SNP) based map in the linkage region of IBD in order to further confine the region where a potential disease susceptibility gene could be located. With 142 SNPs about 20 Mb on the p-arm of chromosome 6 including the MHC were analysed for association within a population of 847 IBD affected individuals and their healthy parents and a group of 550 unrelated blood donors as a control group. Association was tested with a case control and a family based transmission distortion technique (TDT). While three major association leads in the regions of HSP70-HOM to HLA-G, SLC26A8 to MAPK13 and around TREM1 were identified in the case control association, the TDT analysis with Transmit had only one major association lead from HSP70-HOM to rs1035798 (AGER) and 4 minor peaks partly overlapping with the case control results. Based on this, five functional candidate genes were analysed in more detail: HLA-DPA1, MAPK14, MAKP13, BRPF3 and TREM1. In HLA-DPA1 exon 2 polymorphisms at amino acid positions 31, 37-38, and 50 were typed by direct sequencing in affected individuals and if possible healthy family members from Europe and South Africa, and South Korea and unrelated healthy controls from Germany and South Korea. Weak association was only found for CD in the European family population for HLA-DPA1*02021/*02016. Mutation detection in the coding sequence of MAPK14, MAPK13 and TREM1 identified only one exon polymorphism in exon 2 (T25S) of TREM1 and two more

intronic or in the 3' untranslated region. The case control analysis showed positive associations for intronic SNPs in all three genes but not for T25S. From these data neither of these genes is a major determinant of IBD risk. Further analysis of the neighbouring gene BRPF3 did not identify any exonic polymorphisms. Several other functional candidate genes made it plausible that the observed associations were caused through linkage disequilibrium (LD). Extent and structure of LD in the region was unclear. The human genome has been suggested to be organized into haplotype blocks, characterized by high internal levels of LD between SNPs, separated by recombination hot-spots. The distribution of linkage was investigated in a 10 Mb region using 920 SNPs in a population of 45 US-Americans of European ancestry, followed by a more detailed investigation of 3.53 Mb covering the MHC, using 320 SNPs in 550 German, 78 British, 93 Norwegian, and 45 individuals US-Americans of African ancestry. Clustering of Lewontin's D' and correlation between LD in different populations and physical distance between SNPs were implemented. The data show that LD between SNPs can have a variety of characteristics, and the knowledge about this is important for the interpretation of association results. High LD levels and clear haplotype block structures were found to intermingle with areas of low or medium LD, both inside and outside the MHC. All populations with European ancestry show a similar pattern and characteristics of LD structure, while the US-African American populations differs at several areas. These data provide a basis for the analysis and set-up of further association studies in the peri-MHC region, and show the importance of a thorough investigation of LD structures in a region of interest for disease gene detection with association methods.

6. Zusammenfassung: Entwicklung einer SNP-Karte in der erweiterten MHC Region auf Chromosom 6 des Menschen als Instrument zur Identifikation von Kandidatengenen für chronisch entzündliche Darmerkrankungen

Die menschliche MHC Region (Haupt-Histokompatibilitätskomplex) auf Chromosom 6p21 beinhaltet eine Vielzahl an Genen, deren Proteine in der Immunantwort und -regulation von großer Wichtigkeit sind. Dies beinhaltet einerseits die eigentlichen HLA (Humane Leukozytenantigene) Gene, welche für die Erkennung und Präsentation körperfremder Antigene wichtig sind, andererseits aber auch Faktoren der Zytokin-Signaltransduktion, wie zum Beispiel Tumornekrosefaktor-alpha (TNF-α) oder die Mitogen-aktivierten Proteinkinasen (MAPK14). Die Region ist damit ein interessantes Ziel für die Suche nach Genen, die eine Prädisposition für Autoimmunerkrankungen oder chronisch entzündliche Erkrankungen verursachen können. In Kopplungsstudien konnten für verschiedene entzündliche und Autoimmunerkrankungen, wie zum Beispiel chronisch entzündliche Darmerkrankungen (CED), Hinweise auf eine Verbindung mit dem MHC gezeigt werden. Chronisch entzündliche Darmerkrankungen können in zwei Ausprägungen unterteilt werden, Morbus Crohn (MC) und Colitis Ulcerosa (CU). Die Erkrankung ist gekennzeichnet durch wiederkehrende Entzündung des Magen-Darm-Traktes und wird durch ein komplexes Zusammenspiel von genetischen, immunologischen und Umwelt-Faktoren beeinflusst.

Im Rahmen dieser Arbeit wurde zunächst eine auf Punktmutationen (SNPs) basierende Karte entwickelt, die den Bereich der positiven Kopplungsanalyse abdeckt, um ein mögliches Gen mit Prädisposition weiter einzugrenzen. Dazu wurden 142 SNP Marker in einer 20 Mb großen Region, inklusive dem MHC, auf Chromosom 6 in 847 erkrankten Personen sowie deren gesunde Eltern und einer Gruppe von 550 unverwandten Blutspendern als Kontrollgruppe auf Assoziation mit CED untersucht. Die Assoziation wurde in einem Fall-Kontroll-Studienrahmen und mit Hilfe von familienbasiertem Transmissionsungleichgewicht (TDT) getestet. Mit der ersten Methode konnten drei Stellen größerer Assoziation festgestellt werden, HSP70-HOM bis HLA-G, SLC26A8 bis MAPK13 und rund umTREM1. Die zweite Methode hatte eine Hauptassoziation, HSP70-HOM bis rs1035798 (AGER), und mehrere kleinere, die teilweise mit den Ergebnissen aus der Fall-Kontrollstudie überlappten. Auf dieser Grundlage wurden die folgenden fünf funktionellen Kandidatengene ausgewählt und weiter untersucht: HLA-DPA1, MAPK14, MAKP13, BRPF3 und TREM1. In HLA-DPA1

wurden Basenaustausche an den Aminosäurepositionen 31, 37-38 und 50 in Exon 2 untersucht. Beteiligt waren Erkrankte und, soweit vorhanden, gesunde Familienmitglieder aus Europa, Südafrika und Südkorea sowie unverwandte Kontrollpersonen aus Deutschland und Südkorea. Nur in der europäischen Familiengruppe wurde eine schwache Assoziation von MC mit HLA-DPA*02021/*02016 gefunden. Mutationssuche in der kodierenden Sequenz von MAPK14, MAKP13 und TREM1 führte nur zu einem SNP in Exon 2 von TREM1, der einen Aminosäureaustausch zur Folge hatte (T25S), sowie zwei weiteren SNPs im Intron und der 3' untranslatierten Region. Assoziation konnte nur für intronische SNPs der drei Gene in der Fall-Kontrollanalyse nachgewiesen werden. Aufgrund dieser Daten ist keines der oben genannten Gene ein wichtiger Risikofaktor für CED. Ein weiteres benachbartes Gen BRPF3 zeigte ebenfalls keine exonischen SNPs. Die Vielzahl anderer funktioneller Kandidatengene in der Region legt nahe, dass die beobachteten Assoziationen aufgrund von Kopplungsungleichgewicht in der Region entstanden sind. Das Ausmaß, in dem Kopplungsungleichgewicht in der MHC Region vorliegt, ist nicht geklärt. Für das humane Genom wurde allerdings eine Organisation in Haplotypenblöcke mit hoher interner Kopplung, die durch Punkte häufiger Rekombination getrennt sind vorgeschlagen. Die Verteilung der Kopplung wurde zunächst in einer 10 Mb Region mit 920 SNPs in einer Gruppe von 45 US-Amerikanern mit europäischer Abstammung getestet. In der vertieften Analyse einer 3,53 Mb Region, welche den MHC beinhaltet, wurden 320 SNPs zusätzlich in 550 Deutschen, 78 Briten, 93 Norwegern, und 45 US-Amerikanern mit afrikanischer Abstammung getestet. Dazu wurde eine Gruppierung der SNPs aufgrund von Lewontins D' Werten für Kopplung angewendet sowie eine Korrelationsanalyse der Kopplung in den verschiedenen Populationen im Vergleich zur physischen Distanz. Die Daten zeigen, dass Kopplung zwischen SNPs verschiedene Ausprägungen haben kann und dass die Kenntnis darüber wichtig ist für die Auswertung von Assoziationsergebnissen. Klare Haplotyp-Blockstrukturen wurden neben und vermischt mit Regionen mittlerer und geringer Kopplung gefunden, sowohl im MHC als auch außerhalb. Alle Populationen mit europäischer Abstammung zeigten eine ähnliche Struktur und Charakteristik, während die US-Amerikaner afrikanischer Abstammung an einigen Stellen deutliche Unterschiede zeigten. Diese Ergebnisse sind eine wichtige Grundlage für die Analyse und den Versuchsaufbau von Assoziationsstudien im erweiterten MHC Bereich und zeigen die Notwendigkeit gründlicher Kenntnisse der Kopplung in einer Region für die Suche von Krankheitsgenen.

7. References

1. Lamm, L. U., Friedrich, U., Petersen, C. B., Jorgensen, J., Nielsen, J., Therkelsen, A. J. & Kissmeyer-Nielsen, F. (1974) *Hum Hered* **24,** 273-84.

2. van Someren, H., Westerveld, A., Hagemeijer, A., Mees, J. R., Meera Khan, P. & Zaalberg, O. B. (1974) *Proc Natl Acad Sci U S A* **71,** 962-5.

3. consortium, T. M. s. (1999) *Nature* **401,** 921-3.

4. Browning, M. & McMichael, A. (1996) *Oxford: Bios Scientific Publishers*.

5. Stephens, R., Horton, R., Humphray, S., Rowen, L., Trowsdale, J. & Beck, S. (1999) *J Mol Biol* **291,** 789-99.

6. Ruddy, D. A., Kronmal, G. S., Lee, V. K., Mintier, G. A., Quintana, L., Domingo, R., Jr., Meyer, N. C., Irrinki, A., McClelland, E. E., Fullan, A., Mapa, F. A., Moore, T., Thomas, W., Loeb, D. B., Harmon, C., Tsuchihashi, Z., Wolff, R. K., Schatzman, R. C. & Feder, J. N. (1997) *Genome Res* **7,** 441-56.

7. Bodmer, J. G., Marsh, S. G., Albert, E. D., Bodmer, W. F., Bontrop, R. E., Dupont, B., Erlich, H. A., Hansen, J. A., Mach, B., Mayr, W. R., Parham, P., Petersdorf, E. W., Sasazuki, T., Schreuder, G. M., Strominger, J. L., Svejgaard, A. & Terasaki, P. I. (1999) *Hum Immunol* **60,** 361-95.

8. Le Bouteiller, P. (1994) *Crit Rev Immunol* **14,** 89-129.

9. Campbell, R. D., Carroll, M. C. & Porter, R. R. (1986) *Adv Immunol* **38,** 203-44.

10. Campbell, R. D., Dunham, I. & Sargent, C. A. (1988) *Exp Clin Immunogenet* **5,** 81-98.

11. Levine, L. S., Zachmann, M., New, M. I., Prader, A., Pollack, M. S., O'Neill, G. J., Yang, S. Y., Oberfield, S. E. & Dupont, B. (1978) *N Engl J Med* **299,** 911-5.

12. Gruen, J. R. & Weissman, S. M. (1997) *Blood* **90,** 4252-65.

13. Limm, T. M., Ashdown, M. L., Naughton, M. J., McGinnis, M. D. & Simons, M. J. (1993) *Hum Immunol* **38,** 57-68.

14. Horton, R., Niblett, D., Milne, S., Palmer, S., Tubby, B., Trowsdale, J. & Beck, S. (1998) *J Mol Biol* **282,** 71-97.

15. Guillaudeux, T., Janer, M., Wong, G. K., Spies, T. & Geraghty, D. E. (1998) *Proc Natl Acad Sci U S A* **95,** 9494-9.

16. Satta, Y., Kupfermann, H., Li, Y. J. & Takahata, N. (1999) *Immunol Rev* **167,** 367-79.

17. Kaufman, J., Salomonsen, J. & Flajnik, M. (1994) *Semin Immunol* **6,** 411-24.

18. Bontrop, R. E., Otting, N., de Groot, N. G. & Doxiadis, G. G. (1999) *Immunol Rev* **167,** 339-50.

19. Bontrop, R. E., Otting, N., Slierendregt, B. L. & Lanchbury, J. S. (1995) *Immunol Rev* **143,** 33-62.

20. Kasahara, M., Flajnik, M. F., Ishibashi, T. & Natori, T. (1995) *Transpl Immunol* **3,** 1-20.

21. Gaston, J. S., Goodall, J. C., Young, J. L. & Young, S. P. (1997) *Hum Immunol* **54,** 40-7.

22. Begovich, A. B., Moonsamy, P. V., Mack, S. J., Barcellos, L. F., Steiner, L. L., Grams, S., Suraj-Baker, V., Hollenbach, J., Trachtenberg, E., Louie, L., Zimmermann, P., Hill, A. V. S., Stoneking, M., Sasazuki, T., Konenkov, V. I., Sartakova, M. L., Titanji, V. P. K., Rickards, O. & Klitz, W. (2001) *Tissue Antigens* **57,** 424-39.

23. D'Alfonso, S., Bolognesi, E., Mazzola, G., Dall'Omo, A. & Richiardi, P. M. (1999) *J Biol Regul Homeost Agents* **13,** 8-13.

24. Sanchez-Mazas, A., Djoulah, S., Busson, M., Le Monnier de Gouville, I., Poirier, J. C., Dehay, C., Charron, D., Excoffier, L., Schneider, S., Langaney, A., Dausset, J. & Hors, J. (2000) *Eur J Hum Genet* **8,** 33-41.

25. Cullen, M., Perfetto, S. P., Klitz, W., Nelson, G. & Carrington, M. (2002) *Am J Hum Genet* **71,** 759-76.

26. Takahata, N. & Satta, Y. (1998) *Immunogenetics* **47,** 430-41.

27. Perdriger, A., Guggenbuhl, P., Chales, G., Le Dantec, P., Yaouanq, J., Genetet, B., Pawlotsky, Y. & Semana, G. (1996) *Hum Immunol* **46,** 42-8.

28. Veal, C. D., Capon, F., Allen, M. H., Heath, E. K., Evans, J. C., Jones, A., Patel, S., Burden, D., Tillman, D., Barker, J. N. & Trembath, R. C. (2002) *Am J Hum Genet* **71,** 554-64.

29. Schurmann, M., Reichel, P., Muller-Myhsok, B., Schlaak, M., Muller-Quernheim, J. & Schwinger, E. (2001) *Am J Respir Crit Care Med* **164,** 840-6.

30. Hampe, J., Shaw, S. H., Saiz, R., Leysens, N., Lantermann, A., Mascheretti, S., Lynch, N. J., MacPherson, A. J., Bridger, S., van Deventer, S., Stokkers, P., Morin, P., Mirza, M. M., Forbes, A., Lennard-Jones, J. E., Mathew, C. G., Curran, M. E. & Schreiber, S. (1999) *Am J Hum Genet* **65,** 1647-55.

31. Chao, C. C., Sytwu, H. K., Chen, E. L., Toma, J. & McDevitt, H. O. (1999) *Proc Natl Acad Sci U S A* **96,** 9299-304.

32. Pociot, F. & McDermott, M. F. (2002) *Genes Immun* **3,** 235-49.

33. Haines, J. L., Bradford, Y., Garcia, M. E., Reed, A. D., Neumeister, E., Pericak-Vance, M. A., Rimmler, J. B., Menold, M. M., Martin, E. R., Oksenberg, J. R., Barcellos, L. F., Lincoln, R. & Hauser, S. L. (2002) *Hum Mol Genet* **11,** 2251-6.

34. Albig, W., Drabent, B., Burmester, N., Bode, C. & Doenecke, D. (1998) *J Cell Biochem* **69,** 117-26.

35. Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S. L. & Quackenbush, J. (2000) *Nat Genet* **25,** 239-40.

36. Roest Crollius, H., Jaillon, O., Bernot, A., Dasilva, C., Bouneau, L., Fischer, C., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. & Weissenbach, J. (2000) *Nat Genet* **25,** 235-8.

37. Ewing, B. & Green, P. (2000) *Nat Genet* **25,** 232-4.

38. Haldane, J. B. S. & Smith, C. (1947) *Annals of Eugenics* **13,** 10-31.

39. Morton, N. E. (1955) *Am J Hum Genet* **7,** 277-318.

40. Kruglyak, L., Daly, M. J., Reeve-Daly, M. P. & Lander, E. S. (1996) *Am J Hum Genet* **58,** 1347-63.

41. Lander, E. & Kruglyak, L. (1995) *Nat Genet* **11,** 241-7.

42. Spielman, R. S., McGinnis, R. E. & Ewens, W. J. (1993) *Am J Hum Genet* **52,** 506-16.

43. Miller, M. P. (1997) *Northern Arizona University, Flagstaff, Arizona. Freely distributed by the author via the internet.*

44. Spielman, R. S. & Ewens, W. J. (1996) *Am J Hum Genet* **59,** 983-9.

45. Ott, J. (1989) *Genet Epidemiol* **6,** 127-30.

46. Terwilliger, J. D. & Ott, J. (1992) *Hum Hered* **42,** 337-46.

47. Zhao, H., Zhang, S., Merikangas, K. R., Trixler, M., Wildenauer, D. B., Sun, F. & Kidd, K. K. (2000) *Am J Hum Genet* **67,** 936-46.

48. Dawson, E., Abecasis, G. R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D. M., Pabial, J., Dibling, T., Tinsley, E., Kirby, S., Carter, D., Papaspyridonos, M., Livingstone, S., Ganske, R., Lohmussaar, E., Zernant, J., Tonisson, N., Remm, M., Magi, R., Puurand, T., Vilo, J., Kurg, A., Rice, K., Deloukas, P., Mott, R., Metspalu, A., Bentley, D. R., Cardon, L. R. & Dunham, I. (2002) *Nature* **418,** 544-8.

49. Kimura, M. (1983) *Cambridge University Press, New York.*

50. Stephens, J. C., Reich, D. E., Goldstein, D. B., Shin, H. D., Smith, M. W., Carrington, M., Winkler, C., Huttley, G. A., Allikmets, R., Schriml, L., Gerrard, B., Malasky, M., Ramos, M. D., Morlot, S., Tzetis, M., Oddoux, C., di Giovine, F. S., Nasioulas, G.,

Chandler, D., Aseev, M., Hanson, M., Kalaydjieva, L., Glavac, D., Gasparini, P., Dean, M. & et al. (1998) *Am J Hum Genet* **62,** 1507-15.

51. Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. & Lander, E. S. (2001) *Nat Genet* **29,** 229-32.

52. May, C. A., Shone, A. C., Kalaydjieva, L., Sajantila, A. & Jeffreys, A. J. (2002) *Nat Genet* **31,** 272-5.

53. Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J. & Altshuler, D. (2002) *Science* **296,** 2225-9.

54. Reich, D. E., Schaffner, S. F., Daly, M. J., McVean, G., Mullikin, J. C., Higgins, J. M., Richter, D. J., Lander, E. S. & Altshuler, D. (2002) *Nat Genet* **32,** 135-42.

55. Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R. & Lander, E. S. (2001) *Nature* **411,** 199-204.

56. Jeffreys, A. J., Kauppi, L. & Neumann, R. (2001) *Nat Genet* **29,** 217-22.

57. Podolsky, D. K. (1991) *N Engl J Med* **325,** 928-37.

58. Shivananda, S., Lennard-Jones, J., Logan, R., Fear, N., Price, A., Carpenter, L. & van Blankenstein, M. (1996) *Gut* **39,** 690-7.

59. Probert, C. S., Jayanthi, V., Rampton, D. S. & Mayberry, J. F. (1996) *Int J Colorectal Dis* **11,** 25-8.

60. Fiocchi, C. (1998) *Gastroenterology* **115,** 182-205.

61. Munkholm, P. (1997) *Dan Med Bull* **44,** 287-302.

62. Curran, M. E., Lau, K. F., Hampe, J., Schreiber, S., Bridger, S., Macpherson, A. J., Cardon, L. R., Sakul, H., Harris, T. J., Stokkers, P., Van Deventer, S. J., Mirza, M., Raedler, A., Kruis, W., Meckler, U., Theuer, D., Herrmann, T., Gionchetti, P., Lee, J., Mathew, C. & Lennard-Jones, J. (1998) *Gastroenterology* **115,** 1066-71.

63. Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J. P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C. A., Gassull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J. F., Sahbatou, M. & Thomas, G. (2001) *Nature* **411,** 599-603.

64. Ogura, Y., Bonen, D. K., Inohara, N., Nicolae, D. L., Chen, F. F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R. H., Achkar, J. P., Brant, S. R., Bayless, T. M., Kirschner, B. S., Hanauer, S. B., Nunez, G. & Cho, J. H. (2001) *Nature* **411,** 603-6.

65. Hampe, J., Cuthbert, A., Croucher, P. J., Mirza, M., Mascheretti, S., Fisher, S., Frenzel, H., King, K., Hasselmeyer, A., MacPherson, A. J., Bridger, S., Van Deventer, S. J.,

Forbes, A., Nikolaus, S., Lennard-Jones, J., Foelsch, U. R., Krawczak, M., Lewis, C., Schreiber, S. & Mathew, C. (2001) *Lancet* **357,** 1925-8.

66. Schreiber, S. (2000) *Gut* **47,** 746-7.

67. Beutler, B. (2001) *Immunity* **15,** 5-14.

68. Papadakis, K. A. & Targan, S. R. (2000) *Inflamm Bowel Dis* **6,** 303-13.

69. Papadakis, K. A. & Targan, S. R. (2000) *Annu Rev Med* **51,** 289-98.

70. Fuss, I. J., Neurath, M., Boirivant, M., Klein, J. S., de la Motte, C., Strong, S. A., Fiocchi, C. & Strober, W. (1996) *J Immunol* **157,** 1261-70.

71. Plevy, S. E., Landers, C. J., Prehn, J., Carramanzana, N. M., Deem, R. L., Shealy, D. & Targan, S. R. (1997) *J Immunol* **159,** 6276-82.

72. Stallmach, A., Strober, W., MacDonald, T. T., Lochs, H. & Zeitz, M. (1998) *Immunol Today* **19,** 438-41.

73. Murch, S. H., Braegger, C. P., Walker-Smith, J. A. & MacDonald, T. T. (1993) *Gut* **34,** 1705-9.

74. Nikolaus, S., Bauditz, J., Gionchetti, P., Witt, C., Lochs, H. & Schreiber, S. (1998) *Gut* **42,** 470-6.

75. Brandt, E., Colombel, J. F., Ectors, N., Gambiez, L., Emilie, D., Geboes, K., Capron, M. & Desreumaux, P. (2000) *Clin Exp Immunol* **122,** 180-5.

76. Schreiber, S., Nikolaus, S., Hampe, J., Hamling, J., Koop, I., Groessner, B., Lochs, H. & Raedler, A. (1999) *Lancet* **353,** 459-61.

77. Vecchi, M., Bianchi, M. B., Sinico, R. A., Radice, A., Meucci, G., Torgano, G., Omodei, P., Forzenigo, L., Landoni, M., Arrigoni, M. & et al. (1994) *Digestion* **55,** 34-9.

78. Terjung, B., Spengler, U., Sauerbruch, T. & Worman, H. J. (2000) *Gastroenterology* **119,** 310-22.

79. Orholm, M., Munkholm, P., Langholz, E., Nielsen, O. H., Sorensen, I. A. & Binder, V. (1991) *N Engl J Med* **324,** 84-8.

80. Tysk, C., Lindberg, E., Jarnerot, G. & Floderus-Myrhed, B. (1988) *Gut* **29,** 990-6.

81. Thompson, N. P., Driscoll, R., Pounder, R. E. & Wakefield, A. J. (1996) *Bmj* **312,** 95-6.

82. Binder, V. (1998) *Dig Dis* **16,** 351-5.

83. Cho, J. H., Nicolae, D. L., Gold, L. H., Fields, C. T., LaBuda, M. C., Rohal, P. M., Pickles, M. R., Qin, L., Fu, Y., Mann, J. S., Kirschner, B. S., Jabs, E. W., Weber, J.,

Hanauer, S. B., Bayless, T. M. & Brant, S. R. (1998) *Proc Natl Acad Sci U S A* **95,** 7502-7.

84. Rioux, J. D., Daly, M. J., Silverberg, M. S., Lindblad, K., Steinhart, H., Cohen, Z., Delmonte, T., Kocher, K., Miller, K., Guschwan, S., Kulbokas, E. J., O'Leary, S., Winchester, E., Dewar, K., Green, T., Stone, V., Chow, C., Cohen, A., Langelier, D., Lapointe, G., Gaudet, D., Faith, J., Branco, N., Bull, S. B., McLeod, R. S., Griffiths, A. M., Bitton, A., Greenberg, G. R., Lander, E. S., Siminovitch, K. A. & Hudson, T. J. (2001) *Nat Genet* **29,** 223-8.

85. Rioux, J. D., Silverberg, M. S., Daly, M. J., Steinhart, A. H., McLeod, R. S., Griffiths, A. M., Green, T., Brettin, T. S., Stone, V., Bull, S. B., Bitton, A., Williams, C. N., Greenberg, G. R., Cohen, Z., Lander, E. S., Hudson, T. J. & Siminovitch, K. A. (2000) *Am J Hum Genet* **66,** 1863-1870.

86. Satsangi, J., Parkes, M., Louis, E., Hashimoto, L., Kato, N., Welsh, K., Terwilliger, J. D., Lathrop, G. M., Bell, J. I. & Jewell, D. P. (1996) *Nat Genet* **14,** 199-202.

87. Ma, Y., Ohmen, J. D., Li, Z., Bentley, L. G., McElree, C., Pressman, S., Targan, S. R., Fischel-Ghodsian, N., Rotter, J. I. & Yang, H. (1999) *Inflamm Bowel Dis* **5,** 271-8.

88. Duerr, R. H., Barmada, M. M., Zhang, L., Pfutzer, R. & Weeks, D. E. (2000) *Am J Hum Genet* **66,** 1857-62.

89. Hugot, J. P., Laurent-Puig, P., Gower-Rousseau, C., Olson, J. M., Lee, J. C., Beaugerie, L., Naom, I., Dupas, J. L., Van Gossum, A., Orholm, M., Bonaiti-Pellie, C., Weissenbach, J., Mathew, C. G., Lennard-Jones, J. E., Cortot, A., Colombel, J. F. & Thomas, G. (1996) *Nature* **379,** 821-3.

90. Vermeire, S., Satsangi, J., Peeters, M., Parkes, M., Jewell, D. P., Vlietinck, R. & Rutgeerts, P. (2001) *Gastroenterology* **120,** 834-40.

91. Orholm, M., Iselius, L., Sorensen, T. I., Munkholm, P., Langholz, E. & Binder, V. (1993) *Bmj* **306,** 20-4.

92. Hampe, J., Frenzel, H., Mirza, M. M., Croucher, P. J., Cuthbert, A., Mascheretti, S., Huse, K., Platzer, M., Bridger, S., Meyer, B., Nurnberg, P., Stokkers, P., Krawczak, M., Mathew, C. G., Curran, M. & Schreiber, S. (2002) *Proc Natl Acad Sci U S A* **99,** 321-326.

93. Hampe, J., Schreiber, S., Shaw, S. H., Lau, K. F., Bridger, S., Macpherson, A. J., Cardon, L. R., Sakul, H., Harris, T. J., Buckler, A., Hall, J., Stokkers, P., van Deventer, S. J., Nurnberg, P., Mirza, M. M., Lee, J. C., Lennard-Jones, J. E., Mathew, C. G. & Curran, M. E. (1999) *Am J Hum Genet* **64,** 808-16.

94. Hugot, J. P., Laurent-Puig, P., Gower-Rousseau, C., Caillat-Zucman, S., Beaugerie, L., Dupas, J. L., Van Gossum, A., Bonait-Pellie, C., Cortot, A. & Thomas, G. (1994) *Am J Med Genet* **52,** 207-13.

95. Yang, H., Plevy, S. E., Taylor, K., Tyan, D., Fischel-Ghodsian, N., McElree, C., Targan, S. R. & Rotter, J. I. (1999) *Gut* **44,** 519-26.

96. Dechairo, B., Dimon, C., van Heel, D., Mackay, I., Edwards, M., Scambler, P., Jewell, D., Cardon, L., Lench, N. & Carey, A. (2001) *Eur J Hum Genet* **9,** 627-33.

97. Gleeson, M. H., Walker, J. S., Wentzel, J., Chapman, J. A. & Harris, R. (1972) *Gut* **13,** 438-40.

98. van den Berg-Loonen, E. M., Dekker-Saeys, B. J., Meuwissen, S. G., Nijenhuis, L. E. & Engelfriet, C. P. (1977) *J Immunogenet* **4,** 167-75.

99. Biemond, I., Burnham, W. R., D'Amaro, J. & Langman, M. J. (1986) *Gut* **27,** 934-41.

100. Purrmann, J., Korsten, S., Bertrams, J., Miller, B., Lapsien, B., Munch, H., Reis, H. E. & Strohmeyer, G. (1985) *Z Gastroenterol* **23,** 432-7.

101. Fujita, K., Naito, S., Okabe, N. & Yao, T. (1984) *J Clin Lab Immunol* **14,** 99-102.

102. Toyoda, H., Wang, S. J., Yang, H. Y., Redford, A., Magalong, D., Tyan, D., McElree, C. K., Pressman, S. R., Shanahan, F., Targan, S. R. & et al. (1993) *Gastroenterology* **104,** 741-8.

103. Nakajima, A., Matsuhashi, N., Kodama, T., Yazaki, Y., Takazoe, M. & Kimura, A. (1995) *Gastroenterology* **109,** 1462-7.

104. Trachtenberg, E. A., Yang, H., Hayes, E., Vinson, M., Lin, C., Targan, S. R., Tyan, D., Erlich, H. & Rotter, J. I. (2000) *Hum Immunol* **61,** 326-33.

105. Plevy, S. E., Targan, S. R., Yang, H., Fernandez, D., Rotter, J. I. & Toyoda, H. (1996) *Gastroenterology* **110,** 1053-60.

106. Bouma, G., Xia, B., Crusius, J. B., Bioque, G., Koutroubakis, I., Von Blomberg, B. M., Meuwissen, S. G. & Pena, A. S. (1996) *Clin Exp Immunol* **103,** 391-6.

107. Russell, A. S., Percy, J. S., Schlaut, J., Sartor, V. E., Goodhart, J. M., Sherbaniuk, R. W. & Kidd, E. G. (1975) *Am J Dig Dis* **20,** 359-61.

108. Delpre, G., Kadish, U., Gazit, E., Joshua, H. & Zamir, R. (1980) *Gastroenterology* **78,** 1452-7.

109. Smolen, J. S., Gangl, A., Polterauer, P., Menzel, E. J. & Mayr, W. R. (1982) *Gastroenterology* **82,** 34-8.

110. Satsangi, J., Welsh, K. I., Bunce, M., Julier, C., Farrant, J. M., Bell, J. I. & Jewell, D. P. (1996) *Lancet* **347,** 1212-7.

111. Bouma, G., Crusius, J. B., Garcia-Gonzalez, M. A., Meijer, B. U., Hellemans, H. P., Hakvoort, R. J., Schreuder, G. M., Kostense, P. J., Meuwissen, S. G. & Pena, A. S. (1999) *Clin Exp Immunol* **115,** 294-300.

112. Roussomoustakaki, M., Satsangi, J., Welsh, K., Louis, E., Fanning, G., Targan, S., Landers, C. & Jewell, D. P. (1997) *Gastroenterology* **112,** 1845-53.

113. Cariappa, A., Sands, B., Forcione, D., Finkelstein, D., Podolsky, D. K. & Pillai, S. (1998) *Gut* **43,** 210-5.

114. Yoshitake, S., Kimura, A., Okada, M., Yao, T. & Sasazuki, T. (1999) *Tissue Antigens* **53,** 350-8.

115. Stokkers, P. C., Reitsma, P. H., Tytgat, G. N. & van Deventer, S. J. (1999) *Gut* **45,** 395-401.

116. Hoffmeyer, A., Grosse-Wilde, A., Flory, E., Neufeld, B., Kunz, M., Rapp, U. R. & Ludwig, S. (1999) *J Biol Chem* **274,** 4319-27.

117. Waetzig, G. H., Seegert, D., Rosenstiel, P., Nikolaus, S. & Schreiber, S. (2002) *J Immunol* **168,** 5342-51.

118. Enslen, H., Raingeaud, J. & Davis, R. J. (1998) *J Biol Chem* **273,** 1741-8.

119. Lennard-Jones, J. E. (1989) *Scand J Gastroenterol Suppl* **170,** 2-6; discussion 16-9.

120. Hampe, J., Hermann, B., Bridger, S., MacPherson, A. J., Mathew, C. G. & Schreiber, S. (1998) *Int J Colorectal Dis* **13,** 260-3.

121. Curran, M. E., Lau, K. F., Hampe, J., Schreiber, S., Bridger, S., Macpherson, A. J., Cardon, L. R., Sakul, H., Harris, T. J., Stokkers, P., Van Deventer, S. J., Mirza, M., Raedler, A., Kruis, W., Meckler, U., Theuer, D., Herrmann, T., Gionchetti, P., Lee, J., Mathew, C. & Lennard-Jones, J. (1998) *Gastroenterology* **115,** 1066-71.

122. Chomczynski, P., Mackey, K., Drews, R. & Wilfinger, W. (1997) *Biotechniques* **22,** 550-3.

123. Livak, K. J., Flood, S. J., Marmaro, J., Giusti, W. & Deetz, K. (1995) *PCR Methods Appl* **4,** 357-62.

124. Dieffenbach, C. W., Lowe, T. M. & Dveksler, G. S. (1993) *PCR Methods Appl* **3,** S30-7.

125. Livak, K. J., Marmaro, J. & Flood, S. J. (1995) *Research News* **Foster City: Applied Biosystems**.

126. Lowe, T., Sharefkin, J., Yang, S. Q. & Dieffenbach, C. W. (1990) *Nucleic Acids Res* **18,** 1757-61.

127. De La Vega, F. M., Dailey, D., Ziegle, J., Williams, J., Madden, D. & Gilbert, D. A. (2002) *Biotechniques* **Suppl,** 48-50, 52, 54.

128. Mayer, L., Eisenhardt, D., Salomon, P., Bauer, W., Plous, R. & Piccinini, L. (1991) *Gastroenterology* **100,** 3-12.

129. Momburg, F., Koretz, K., Von Herbay, A. & Moller, P. (1988) *Clin Exp Immunol* **72,** 367-72.

130. Horie, Y., Chiba, M., Iizuka, M. & Masamune, O. (1990) *Gastroenterol Jpn* **25,** 575-84.

131. Koretz, K., Momburg, F., Otto, H. F. & Moller, P. (1987) *Am J Pathol* **129,** 493-502.

132. Sturgess, R. P., Hooper, L. B., Spencer, J., Hung, C. H., Nelufer, J. M. & Ciclitira, P. J. (1992) *Scand J Gastroenterol* **27,** 907-11.

133. Salomon, P., Pizzimenti, A., Panja, A., Reisman, A. & Mayer, L. (1991) *Autoimmunity* **9,** 141-9.

134. Pallone, F., Fais, S. & Capobianchi, M. R. (1988) *Clin Exp Immunol* **74,** 75-9.

135. Nagy, M. & Roewer, L. (1992) *Ann Univ Sarav Med* **Suppl. 9,** 145-152.

136. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239,** 487-91.

137. Williams, J. F. (1989) *Biotechniques* **7,** 762-9.

138. Dracopoli, N. C., L., H. J., R., K. B., T., M. D., .C., M. C., E., S. C., G., S. J. & R., S. D. (1994) *John Wiley & Sons Ltd*.

139. Herbison, C. E., Sayer, D. C., Bellgard, M., Allcock, R. J., Christiansen, F. T. & Price, P. (1999) *DNA Seq* **10,** 229-43.

140. Faccio, L., Chen, A., Fusco, C., Martinotti, S., Bonventre, J. V. & Zervos, A. S. (2000) *Am J Physiol Cell Physiol* **278,** C781-90.

141. Zervos, A. S., Faccio, L., Gatto, J. P., Kyriakis, J. M. & Brent, R. (1995) *Proc Natl Acad Sci U S A* **92,** 10531-4.

142. Han, J., Richter, B., Li, Z., Kravchenko, V. & Ulevitch, R. J. (1995) *Biochim Biophys Acta* **1265,** 224-7.

143. Lee, J. C., Laydon, J. T., McDonnell, P. C., Gallagher, T. F., Kumar, S., Green, D., McNulty, D., Blumenthal, M. J., Heys, J. R., Landvatter, S. W. & et al. (1994) *Nature* **372,** 739-46.

144. Goedert, M., Cuenda, A., Craxton, M., Jakes, R. & Cohen, P. (1997) *Embo J* **16,** 3563-71.

145. Bouchon, A., Dietrich, J. & Colonna, M. (2000) *J Immunol* **164,** 4991-5.

146. Hampe, J., Wollstein, A., Lu, T., Frevel, H. J., Will, M., Manaster, C. & Schreiber, S. (2001) *Bioinformatics* **17,** 654-5.

147. Hall, J. & Nanthakumar, E. (1997) *In: Boyle AL (ed) Current protocols in human genetics. Vol 2. John Wiley & Sons, New York, pp 2.8.1- 2.8.19*.

148. Clayton, D. (1999) *Am J Hum Genet* **65,** 1170-7.

149. Lewontin, R. C. (1964) *Genetics* **49,** 49-67.

150. Lewontin, R. C. (1988) *Genetics* **120,** 849-52.

151. Terwiliger, J. & Ott, J. (1994) *Handbook of Human Genetic Linkage* (Johns Hopkins University Press, Baltimore).

152. Schneider, S., Roessli, D. & Excoffier, L. (2000) *Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva*.

153. Wright, S. (1969) *University of Chicago Press* **2,** 295.

154. Durbin, R., Eddy, S., Krogh, A. & Mitchison, G. (1998) *Cambridge University Press, Cambridge***,** pp 166-169.

155. Maniatis, N., Collins, A., Xu, C. F., McCarthy, L. C., Hewett, D. R., Tapper, W., Ennis, S., Ke, X. & Morton, N. E. (2002) *Proc Natl Acad Sci U S A* **99,** 2228-33.

156. Morton, N. E., Zhang, W., Taillon-Miller, P., Ennis, S., Kwok, P. Y. & Collins, A. (2001) *Proc Natl Acad Sci U S A* **98,** 5217-21.

157. (1996) *Hum Immunol* **47,** 1-184.

158. Martin, E. R., Lai, E. H., Gilbert, J. R., Rogala, A. R., Afshari, A. J., Riley, J., Finch, K. L., Stevens, J. F., Livak, K. J., Slotterbeck, B. D., Slifer, S. H., Warren, L. L., Conneally, P. M., Schmechel, D. E., Purvis, I., Pericak-Vance, M. A., Roses, A. D. & Vance, J. M. (2000) *Am J Hum Genet* **67,** 383-94.

159. Weiss, K. M. & Terwilliger, J. D. (2000) *Nat Genet* **26,** 151-7.

160. Ohashi, J. & Tokunaga, K. (2001) *J Hum Genet* **46,** 478-82.

161. Wright, J. P., Froggatt, J., O'Keefe, E. A., Ackerman, S., Watermeyer, S., Louw, J., Adams, G., Girdwood, A. H., Burns, D. G. & Marks, I. N. (1986) *S Afr Med J* **70,** 10-5.

162. Waetzig, G. H., Seegert, D., Nikolaus, S., Rosenstiel, P., Sfikas, N. & Schreiber, S. (2001) *Gastroenterology* **120,** A522.

163. Hommes, D., van den Blink, B., Plasse, T., Bartelsman, J., Xu, C., Macpherson, B., Tytgat, G., Peppelenbosch, M. & Van Deventer, S. (2002) *Gastroenterology* **122,** 7-14.

164. Kultz, D. (1998) *J Mol Evol* **46,** 571-88.

165. Cuthbert, A. P., Fisher, S. A., Mirza, M. M., King, K., Hampe, J., Croucher, P. J., Mascheretti, S., Sanderson, J., Forbes, A., Mansfield, J., Schreiber, S., Lewis, C. M. & Mathew, C. G. (2002) *Gastroenterology* **122,** 867-74.

166. Ahmad, T., Armuzzi, A., Bunce, M., Mulcahy-Hawes, K., Marshall, S. E., Orchard, T. R., Crawshaw, J., Large, O., de Silva, A., Cook, J. T., Barnardo, M., Cullen, S., Welsh, K. I. & Jewell, D. P. (2002) *Gastroenterology* **122,** 854-66.

167. Orchard, T. R., Thiyagaraja, S., Welsh, K. I., Wordsworth, B. P., Hill Gaston, J. S. & Jewell, D. P. (2000) *Gastroenterology* **118,** 274-8.

168. Frisse, L., Hudson, R. R., Bartoszewicz, A., Wall, J. D., Donfack, J. & Di Rienzo, A. (2001) *Am J Hum Genet* **69,** 831-43.

169. Lercher, M. J. & Hurst, L. D. (2002) *Trends Genet* **18,** 337-40.

170. Cereb, N., Hughes, A. L. & Yang, S. Y. (1997) *Immunogenetics* **47,** 30-6.

171. Meyer, D. & Thomson, G. (2001) *Ann Hum Genet* **65,** 1-26.

172. De La Vega, F. M., Su, X., Avi-Itzhak, H. I., Halldorson, B. V., Gordon, D., Collins, A., Lippert, R., Schwartz, R., Scafe, C. R., Wang, Y., Laig-Webster, M., Koehler, R. T., Ziegle, J. S., Wogan, L. T., Stevens, J. F., Leinen, K. M., Olson, S. J., Guegler, K. J., You, X., Xu, L., Hemken, H. G., Kalush, F., Clark, A. G., Istrail, S., Hunkapiller, M. W., Spier, E. G. & Gilbert, D. A. (2003) *submitted.*

173. Jeffreys, A. J. & Neumann, R. (2002) *Nat Genet* **31,** 267-71.

174. Guillon, H. & de Massy, B. (2002) *Nat Genet* **32,** 296-9.

175. Jeffreys, A. J., Murray, J. & Neumann, R. (1998) *Mol Cell* **2,** 267-73.

8. Appendix

8.1. Supplemental data

8.1.1. Allele and genotype association analysis

Table 8.1 part1: Allele and genotype frequencies in 142 SNP markers

		Allele frequencies								Genotype frequencies											
		Controls		IBD cases		CD cases		UC cases		Controls			IBD cases			CD cases			UC cases		
SNP marker name	Position (Mb)	1	2	1	2	1	2	1	2	1-1	1-2	2-2	1-1	1-2	2-2	1-1	1-2	2-2	1-1	1-2	2-2
rs1563727	21.29	0.37	0.63	0.39	0.61	0.40	0.60	0.38	0.62	0.14	0.46	0.40	0.15	0.48	0.37	0.16	0.47	0.37	0.13	0.51	0.37
rs1935005	22.25	0.78	0.22	0.80	0.20	0.79	0.21	0.81	0.19	0.61	0.34	0.05	0.63	0.33	0.04	0.62	0.33	0.04	0.65	0.32	0.03
rs714591	22.42	0.56	0.44	0.56	0.44	0.55	0.45	0.58	0.42	0.32	0.49	0.19	0.32	0.48	0.20	0.32	0.46	0.21	0.32	0.51	0.17
rs867543	22.58	0.57	0.43	0.55	0.45	0.55	0.45	0.55	0.45	0.33	0.49	0.18	0.30	0.49	0.21	0.29	0.50	0.20	0.32	0.46	0.22
rs1997781	22.74	0.78	0.22	0.77	0.23	0.77	0.23	0.79	0.21	0.62	0.33	0.05	0.61	0.33	0.06	0.59	0.36	0.05	0.65	0.28	0.07
rs1024216	22.77	0.58	0.42	0.55	0.45	0.55	0.45	0.55	0.45	0.34	0.49	0.17	0.30	0.51	0.19	0.28	0.54	0.18	0.31	0.48	0.21
rs1928083	22.96	0.64	0.36	0.65	0.35	0.65	0.35	0.63	0.37	0.42	0.44	0.13	0.43	0.44	0.13	0.44	0.43	0.13	0.40	0.47	0.14
rs1325016	23.81	0.20	0.80	0.20	0.80	0.21	0.79	0.17	0.83	0.03	0.33	0.64	0.03	0.32	0.64	0.04	0.33	0.63	0.02	0.29	0.68
rs1402405	24.15	0.59	0.41	0.56	0.44	0.56	0.44	0.55	0.45	0.35	0.47	0.18	0.32	0.48	0.20	0.31	0.50	0.19	0.32	0.45	0.23
rs932494	24.97	0.61	0.39	0.60	0.40	0.60	0.40	0.59	0.41	0.41	0.40	0.19	0.36	0.48	0.16	0.37	0.47	0.16	0.34	0.51	0.16
rs302976	25.18	0.19	0.81	0.18	0.82	0.17	0.83	0.19	0.81	0.04	0.30	0.66	0.02	0.31	0.67	0.02	0.31	0.68	0.03	0.31	0.66
rs1318016	25.82	0.77	0.23	0.73	0.27	0.74	0.26	0.72	0.28	0.59	0.35	0.06	0.54	0.40	0.07	0.53	0.41	0.06	0.53	0.38	0.09
rs1141034	25.84	0.25	0.75	0.28	0.72	0.28	0.72	0.28	0.72	0.06	0.39	0.56	0.07	0.41	0.52	0.06	0.44	0.50	0.09	0.38	0.52
rs1747522	25.88	0.50	0.50	0.49	0.51	0.49	0.51	0.48	0.52	0.24	0.52	0.25	0.24	0.50	0.27	0.24	0.50	0.26	0.23	0.49	0.28
rs1540273	25.98	0.37	0.63	0.38	0.62	0.38	0.62	0.38	0.62	0.13	0.48	0.39	0.15	0.47	0.38	0.14	0.49	0.38	0.17	0.44	0.40
rs1436310	26.03	0.60	0.40	0.58	0.42	0.58	0.42	0.58	0.42	0.37	0.46	0.18	0.33	0.49	0.18	0.33	0.50	0.17	0.34	0.48	0.18
rs1535276	26.58	0.46	0.54	0.45	0.55	0.45	0.55	0.43	0.57	0.20	0.52	0.28	0.20	0.51	0.29	0.19	0.51	0.29	0.18	0.50	0.31
rs126007	27.80	0.23	0.77	0.23	0.77	0.25	0.75	0.21	0.79	0.06	0.34	0.60	0.06	0.36	0.59	0.06	0.37	0.57	0.04	0.35	0.61
rs126136	28.35	0.71	0.29	0.71	0.29	0.70	0.30	0.72	0.28	0.53	0.37	0.10	0.51	0.41	0.09	0.50	0.40	0.10	0.50	0.44	0.06
rs209179	28.85	0.51	0.49	0.45	0.55	0.46	0.54	0.46	0.54	0.25	0.51	0.24	0.21	0.50	0.30	0.22	0.47	0.31	0.18	0.56	0.26
rs929042	28.95	0.77	0.23	0.81	0.19	0.80	0.20	0.81	0.19	0.59	0.36	0.05	0.64	0.33	0.03	0.65	0.32	0.04	0.63	0.34	0.02
rs1362073	29.44	0.30	0.70	0.34	0.66	0.33	0.67	0.36	0.64	0.09	0.42	0.49	0.12	0.44	0.44	0.11	0.44	0.45	0.14	0.43	0.42
rs404240	29.58	0.97	0.03	0.97	0.03	0.97	0.03	0.97	0.03	0.93	0.07	0.00	0.95	0.05		0.94	0.05		0.95	0.05	0.00
rs362536	29.58	0.45	0.55	0.52	0.48	0.50	0.50	0.54	0.46	0.20	0.51	0.29	0.26	0.52	0.22	0.23	0.55	0.22	0.30	0.46	0.23
rs64036	29.58	0.97	0.03	0.97	0.03	0.98	0.02	0.97	0.03	0.94	0.06		0.95	0.05		0.95	0.05		0.95	0.05	
GABABR1_E7_G141esA	29.64	0.90	0.10	0.89	0.11	0.88	0.12	0.90	0.10	0.80	0.20		0.79	0.20	0.01	0.77	0.22	0.01	0.81	0.18	0.01
MOG_E3_G550A	29.69	0.06	0.94	0.12	0.88	0.11	0.89	0.13	0.87		0.12	0.88	0.01	0.21	0.78	0.01	0.20	0.79	0.02	0.22	0.77
rs1476572	29.73	0.78	0.22	0.71	0.29	0.72	0.28	0.70	0.30	0.62	0.31	0.07	0.51	0.40	0.08	0.52	0.40	0.08	0.50	0.40	0.10
rs1569315	29.85	0.04	0.96	0.05	0.95	0.05	0.95	0.03	0.97		0.09	0.91		0.09	0.91		0.10	0.90		0.07	0.93
HCG-V_3pu_T3056C	30.08	0.15	0.85	0.13	0.87	0.12	0.88	0.13	0.87	0.02	0.28	0.71	0.02	0.21	0.77	0.02	0.19	0.78	0.01	0.23	0.75
HLA-E_E3_A1458G	30.51	0.44	0.56	0.47	0.53	0.47	0.53	0.47	0.53	0.17	0.54	0.29	0.22	0.49	0.29	0.22	0.50	0.28	0.24	0.47	0.29
GTF2H4G-24T	30.93	0.20	0.80	0.20	0.80	0.20	0.80	0.19	0.81	0.05	0.30	0.66	0.04	0.32	0.65	0.04	0.32	0.64	0.04	0.31	0.65
CDSN_E1_A66T	31.14	0.20	0.80	0.18	0.82	0.20	0.80	0.15	0.85	0.04	0.32	0.64	0.03	0.30	0.67	0.03	0.34	0.63	0.03	0.25	0.72
rs1265109	31.18	0.25	0.75	0.23	0.77	0.23	0.77	0.21	0.79	0.09	0.33	0.58	0.05	0.35	0.60	0.05	0.38	0.58	0.06	0.31	0.63
SC1_E2_G2712A	31.19	0.26	0.74	0.29	0.71	0.26	0.74	0.34	0.66	0.07	0.37	0.56	0.10	0.38	0.53	0.07	0.38	0.55	0.13	0.41	0.45
HLA-C_E2_G1019A	31.29	0.39	0.61	0.37	0.63	0.38	0.62	0.35	0.65	0.14	0.49	0.37	0.13	0.48	0.39	0.13	0.49	0.38	0.12	0.47	0.41
HLA-B_E5_G582A	31.38	0.36	0.64	0.35	0.65	0.35	0.65	0.35	0.65	0.12	0.48	0.41	0.13	0.45	0.43	0.13	0.43	0.44	0.11	0.47	0.41
rs2005423	31.41	0.82	0.18	0.84	0.16	0.85	0.15	0.84	0.16	0.66	0.33	0.01	0.72	0.25	0.03	0.72	0.25	0.03	0.72	0.24	0.04
MICA_E4_A8561G	31.43	0.58	0.42	0.57	0.43	0.56	0.44	0.58	0.42	0.32	0.52	0.16	0.33	0.49	0.18	0.32	0.48	0.20	0.33	0.51	0.16
MICB_E3_C7991G	31.53	0.13	0.87	0.12	0.88	0.13	0.87	0.11	0.89	0.01	0.23	0.75	0.01	0.21	0.78	0.02	0.23	0.76	0.02	0.18	0.80
TNFA_G308A_	31.61	0.18	0.82	0.17	0.83	0.17	0.83	0.17	0.83	0.03	0.31	0.66	0.03	0.28	0.69	0.03	0.29	0.69	0.03	0.28	0.69
rs707916	31.76	0.60	0.40	0.63	0.37	0.61	0.39	0.67	0.33	0.37	0.46	0.16	0.41	0.44	0.15	0.38	0.45	0.17	0.47	0.42	0.12
rs707926	31.82	0.16	0.84	0.16	0.84	0.18	0.82	0.13	0.87	0.02	0.28	0.70	0.03	0.25	0.72	0.04	0.27	0.69	0.03	0.20	0.77
HSP70-HOM_G2763A	31.85	0.34	0.66	0.31	0.69	0.29	0.71	0.34	0.66	0.13	0.42	0.45	0.10	0.43	0.47	0.09	0.40	0.51	0.10	0.48	0.42
rs539689	31.87	0.48	0.52	0.46	0.54	0.45	0.55	0.45	0.55	0.22	0.53	0.25	0.22	0.48	0.30	0.23	0.45	0.32	0.18	0.54	0.28
C2_E7_G954C	31.97	0.04	0.96	0.03	0.97	0.03	0.97	0.04	0.96		0.08	0.92		0.07	0.93		0.06	0.94		0.07	0.93
rs1035798	32.18	0.70	0.30	0.74	0.26	0.76	0.24	0.71	0.29	0.49	0.42	0.09	0.56	0.35	0.08	0.60	0.33	0.08	0.52	0.39	0.09
BTNL-II_E3_A158G	32.40	0.36	0.64	0.31	0.69	0.33	0.67	0.29	0.71	0.09	0.53	0.38	0.10	0.43	0.47	0.11	0.45	0.45	0.09	0.41	0.50
HLA-DRA_E4_T724G	32.44	0.57	0.43	0.57	0.43	0.56	0.44	0.56	0.44	0.33	0.49	0.18	0.32	0.50	0.18	0.30	0.52	0.18	0.31	0.49	0.19
HLA-DQA1_E3_C389T	32.63	0.29	0.71	0.25	0.75	0.25	0.75	0.26	0.74	0.08	0.42	0.51	0.08	0.36	0.57	0.06	0.38	0.56	0.10	0.31	0.58
TAP2_E9_G1693A	32.79	0.07	0.93	0.06	0.94	0.07	0.93	0.06	0.94		0.13	0.86		0.12	0.88		0.12	0.87		0.11	0.89
HLA-DMB_E3_T590C	32.90	0.16	0.84	0.20	0.80	0.18	0.82	0.21	0.79	0.04	0.23	0.73	0.04	0.31	0.65	0.03	0.30	0.67	0.04	0.32	0.63
rs1063478	32.91	0.86	0.14	0.86	0.14	0.85	0.15	0.86	0.14	0.74	0.23	0.02	0.75	0.22	0.02	0.73	0.24	0.03	0.76	0.20	0.03
rs206775	32.95	0.59	0.41	0.60	0.40	0.60	0.40	0.60	0.40	0.36	0.46	0.19	0.36	0.49	0.16	0.35	0.50	0.15	0.37	0.45	0.18
RING3_E8_C6193T	32.96	0.04	0.96	0.03	0.97	0.04	0.96	0.03	0.97		0.08	0.92		0.07	0.93		0.08	0.92		0.05	0.95
rs2581	32.98	0.45	0.55	0.43	0.57	0.43	0.57	0.42	0.58	0.20	0.50	0.31	0.18	0.49	0.33	0.17	0.50	0.32	0.18	0.48	0.34
rs926424	33.15	0.29	0.71	0.30	0.70	0.29	0.71	0.34	0.66	0.09	0.40	0.51	0.09	0.43	0.48	0.09	0.40	0.51	0.10	0.47	0.43
rs1547387	33.16	0.12	0.88	0.12	0.88	0.12	0.88	0.13	0.87	0.02	0.21	0.77	0.01	0.23	0.76	0.01	0.23	0.76		0.24	0.75
rs466384	33.24	0.12	0.88	0.11	0.89	0.12	0.88	0.11	0.89	0.01	0.23	0.77	0.01	0.21	0.78	0.01	0.21	0.78	0.01	0.19	0.80
TAPBP_E4_G325C	33.27	0.46	0.54	0.48	0.52	0.48	0.52	0.49	0.51	0.21	0.51	0.28	0.24	0.48	0.28	0.24	0.48	0.28	0.26	0.46	0.28
rs999942	33.62	0.75	0.25	0.75	0.25	0.74	0.26	0.76	0.24	0.58	0.34	0.08	0.56	0.37	0.06	0.54	0.40	0.06	0.59	0.35	0.06
rs10672	33.66	0.32	0.68	0.33	0.67	0.32	0.68	0.36	0.64	0.11	0.41	0.48	0.11	0.44	0.45	0.10	0.44	0.46	0.13	0.45	0.42
rs747694	33.73	0.45	0.55	0.48	0.52	0.47	0.53	0.49	0.51	0.19	0.52	0.29	0.23	0.50	0.27	0.23	0.49	0.28	0.22	0.54	0.24
rs943461	33.77	0.77	0.23	0.78	0.22	0.78	0.22	0.79	0.21	0.61	0.34	0.06	0.61	0.34	0.05	0.60	0.35	0.04	0.63	0.32	0.05
rs902196	34.00	0.02	0.98	0.04	0.96	0.04	0.96	0.03	0.97		0.04	0.96		0.08	0.92		0.09	0.91		0.05	0.95
rs914815	34.12	0.54	0.46	0.52	0.48	0.51	0.49	0.56	0.44	0.29	0.51	0.20	0.27	0.51	0.22	0.25	0.51	0.24	0.30	0.52	0.18

Table 8.1 part2: Allele and genotype frequencies

		Allele frequencies								Genotype frequencies											
	Position	Controls		IBD cases		CD cases		UC cases		Controls			IBD cases			CD cases			UC cases		
SNP marker name	(Mb)	1	2	1	2	1	2	1	2	1-1	1-2	2-2	1-1	1-2	2-2	1-1	1-2	2-2	1-1	1-2	2-2
rs1759627	34.19	0.17	0.83	0.17	0.83	0.16	0.84	0.17	0.83	0.04	0.27	0.69	0.02	0.29	0.69	0.02	0.28	0.70	0.02	0.30	0.68
hCV2489254	35.42	0.30	0.70	0.33	0.67	0.32	0.68	0.34	0.66	0.08	0.43	0.49	0.13	0.40	0.46	0.13	0.40	0.48	0.14	0.41	0.45
hCV2489242	35.44	0.62	0.38	0.58	0.42	0.59	0.41	0.57	0.43	0.38	0.49	0.13	0.36	0.44	0.20	0.37	0.44	0.19	0.36	0.43	0.22
rs755658	35.55	0.90	0.10	0.92	0.08	0.92	0.08	0.91	0.09	0.81	0.18	0.01	0.84	0.15	0.01	0.84	0.14	0.01	0.83	0.16	
rs992105	35.55	0.83	0.17	0.83	0.17	0.83	0.17	0.82	0.18	0.68	0.30	0.02	0.68	0.29	0.03	0.69	0.29	0.03	0.67	0.30	0.03
rs1320991	35.56	0.21	0.79	0.22	0.78	0.22	0.78	0.22	0.78	0.03	0.36	0.61	0.05	0.34	0.61	0.05	0.34	0.61	0.06	0.32	0.62
hCV1256778	35.57	0.27	0.73	0.28	0.72	0.28	0.72	0.31	0.69	0.07	0.42	0.52	0.08	0.41	0.51	0.07	0.40	0.52	0.10	0.43	0.48
rs1360780	35.60	0.68	0.32	0.69	0.31	0.69	0.31	0.69	0.31	0.46	0.45	0.09	0.47	0.43	0.09	0.48	0.42	0.10	0.46	0.46	0.08
hCV15812414	35.70	0.83	0.17	0.82	0.18	0.82	0.18	0.83	0.17	0.68	0.30	0.02	0.67	0.30	0.03	0.66	0.31	0.03	0.68	0.28	0.03
hCV1979181	35.73	0.40	0.60	0.39	0.61	0.40	0.60	0.37	0.63	0.14	0.51	0.35	0.16	0.46	0.38	0.15	0.49	0.36	0.17	0.41	0.42
rs1326752	35.83	0.69	0.31	0.67	0.33	0.67	0.33	0.69	0.31	0.45	0.46	0.08	0.45	0.44	0.10	0.44	0.45	0.11	0.47	0.44	0.09
hCV2472420	35.85	0.69	0.31	0.68	0.32	0.67	0.33	0.69	0.31	0.46	0.45	0.08	0.46	0.44	0.10	0.46	0.44	0.11	0.47	0.44	0.09
hCV617976	35.88	0.69	0.31	0.68	0.32	0.67	0.33	0.69	0.31	0.46	0.46	0.08	0.46	0.44	0.10	0.45	0.44	0.11	0.47	0.44	0.09
hCV617970	35.89	0.66	0.34	0.64	0.36	0.64	0.36	0.65	0.35	0.42	0.49	0.10	0.41	0.47	0.13	0.40	0.47	0.13	0.42	0.48	0.11
hCV3001847	35.96	0.42	0.58	0.47	0.53	0.48	0.52	0.46	0.54	0.17	0.51	0.32	0.22	0.50	0.28	0.23	0.50	0.27	0.21	0.49	0.29
hCV3001832	35.98	0.76	0.24	0.76	0.24	0.76	0.24	0.75	0.25	0.58	0.36	0.06	0.58	0.35	0.06	0.59	0.35	0.06	0.56	0.37	0.07
hCV8852296	35.98	0.34	0.66	0.29	0.71	0.28	0.72	0.27	0.73	0.10	0.46	0.43	0.08	0.42	0.50	0.07	0.42	0.51	0.08	0.38	0.54
hCV3001824	36.00	0.53	0.47	0.46	0.54	0.45	0.55	0.48	0.52	0.27	0.51	0.22	0.22	0.49	0.30	0.20	0.50	0.31	0.24	0.47	0.29
hCV8852251	36.01	0.17	0.83	0.15	0.85	0.14	0.86	0.17	0.83	0.04	0.26	0.70	0.02	0.26	0.71	0.02	0.25	0.74	0.03	0.29	0.69
hCV2455932	36.06	0.12	0.88	0.09	0.91	0.10	0.90	0.07	0.93		0.23	0.76	0.01	0.16	0.83	0.01	0.17	0.82	0.01	0.12	0.87
hCV2455926	36.07	0.54	0.46	0.47	0.53	0.46	0.54	0.48	0.52	0.29	0.50	0.21	0.23	0.47	0.30	0.21	0.49	0.30	0.25	0.45	0.30
hCV2455924	36.07	0.47	0.53	0.53	0.47	0.54	0.46	0.52	0.48	0.22	0.50	0.28	0.29	0.48	0.23	0.30	0.49	0.21	0.30	0.46	0.25
hCV2455922	36.08	0.27	0.73	0.32	0.68	0.33	0.67	0.31	0.69	0.07	0.39	0.53	0.11	0.42	0.47	0.11	0.43	0.46	0.11	0.41	0.48
hCV2455919	36.08	0.80	0.20	0.83	0.17	0.84	0.16	0.81	0.19	0.65	0.31	0.05	0.68	0.29	0.03	0.70	0.27	0.03	0.66	0.30	0.04
rs763096	36.10	0.66	0.34	0.60	0.40	0.59	0.41	0.59	0.41	0.44	0.45	0.11	0.34	0.52	0.14	0.34	0.50	0.16	0.33	0.53	0.14
rs2007683	36.10	0.36	0.64	0.34	0.66	0.33	0.67	0.33	0.67	0.13	0.46	0.41	0.10	0.47	0.43	0.11	0.45	0.44	0.09	0.48	0.43
rs2815812	36.10	0.30	0.70	0.26	0.74	0.26	0.74	0.25	0.75	0.09	0.41	0.49	0.07	0.38	0.55	0.07	0.38	0.55	0.07	0.35	0.57
rs2071863	36.10	0.22	0.78	0.17	0.83	0.17	0.83	0.16	0.84	0.04	0.38	0.59	0.03	0.27	0.70	0.03	0.27	0.70	0.04	0.25	0.71
hCV2476692	36.24	0.71	0.29	0.68	0.32	0.66	0.34	0.73	0.27	0.51	0.40	0.09	0.46	0.43	0.10	0.44	0.44	0.12	0.53	0.39	0.08
hCV2476745	36.27	0.71	0.29	0.71	0.29	0.70	0.30	0.71	0.29	0.53	0.37	0.10	0.50	0.41	0.09	0.51	0.39	0.10	0.47	0.47	0.06
rs744848	36.33	0.19	0.81	0.18	0.82	0.17	0.83	0.19	0.81	0.04	0.30	0.66	0.03	0.30	0.68	0.03	0.29	0.69	0.04	0.31	0.65
rs763017	36.33	0.01	0.99	0.01	0.99	0.01	0.99	0.02	0.98		0.02	0.98		0.03	0.97		0.03	0.97		0.03	0.97
hCV3247999	36.34	0.86	0.14	0.85	0.15	0.86	0.14	0.82	0.18	0.75	0.23	0.02	0.71	0.27	0.02	0.74	0.25	0.02	0.66	0.31	0.02
rs881648	36.35	0.12	0.88	0.14	0.86	0.14	0.86	0.14	0.86	0.01	0.22	0.77	0.02	0.24	0.74	0.02	0.24	0.74	0.03	0.23	0.75
rs881647	36.35	0.40	0.60	0.41	0.59	0.41	0.59	0.41	0.59	0.15	0.49	0.36	0.17	0.47	0.36	0.18	0.46	0.36	0.16	0.49	0.34
rs1885205	36.35	0.88	0.12	0.87	0.13	0.87	0.13	0.86	0.14	0.78	0.20	0.02	0.75	0.23	0.02	0.76	0.22	0.02	0.74	0.23	0.03
hCV16175628	36.44	0.19	0.81	0.19	0.81	0.19	0.81	0.19	0.81	0.04	0.29	0.67	0.04	0.30	0.66	0.04	0.29	0.67	0.03	0.32	0.65
hCV7516771	36.46	0.19	0.81	0.19	0.81	0.19	0.81	0.19	0.81	0.04	0.30	0.66	0.04	0.30	0.66	0.04	0.29	0.67	0.03	0.32	0.65
hCV791696	36.47	0.70	0.30	0.70	0.30	0.71	0.29	0.68	0.32	0.52	0.36	0.12	0.50	0.42	0.09	0.52	0.38	0.10	0.45	0.46	0.09
hCV3053930	36.49	0.27	0.73	0.27	0.73	0.26	0.74	0.28	0.72	0.09	0.37	0.54	0.07	0.39	0.54	0.08	0.36	0.56	0.07	0.44	0.50
hCV2969554	36.56	0.80	0.20	0.79	0.21	0.79	0.21	0.77	0.23	0.64	0.31	0.05	0.62	0.34	0.04	0.63	0.33	0.04	0.59	0.37	0.04
rs1059234	36.65	0.92	0.08	0.94	0.06	0.94	0.06	0.93	0.07	0.84	0.16		0.88	0.12		0.89	0.11		0.87	0.13	
hCV2969521	36.66	0.42	0.58	0.44	0.56	0.43	0.57	0.45	0.55	0.18	0.49	0.33	0.20	0.48	0.32	0.19	0.49	0.32	0.23	0.44	0.33
hCV1022631	36.69	0.40	0.60	0.39	0.61	0.40	0.60	0.38	0.62	0.18	0.44	0.38	0.16	0.47	0.37	0.16	0.48	0.37	0.16	0.45	0.39
hCV1022677	36.73	0.64	0.36	0.65	0.35	0.64	0.36	0.65	0.35	0.39	0.50	0.12	0.43	0.44	0.13	0.42	0.45	0.13	0.43	0.44	0.13
hCV3211720	36.81	0.34	0.66	0.35	0.65	0.35	0.65	0.36	0.64	0.14	0.41	0.45	0.13	0.45	0.42	0.11	0.47	0.42	0.14	0.43	0.43
hCV2061156	36.84	0.73	0.27	0.78	0.22	0.79	0.21	0.76	0.24	0.53	0.40	0.07	0.61	0.34	0.05	0.61	0.36	0.04	0.59	0.34	0.07
hCV2061161	36.84	0.56	0.44	0.53	0.47	0.53	0.47	0.52	0.48	0.33	0.46	0.21	0.29	0.49	0.23	0.28	0.50	0.22	0.29	0.48	0.24
hCV2703063	36.93	0.45	0.55	0.47	0.53	0.45	0.55	0.48	0.52	0.21	0.49	0.30	0.23	0.48	0.29	0.21	0.48	0.31	0.25	0.47	0.28
hCV2317531	36.95	0.17	0.83	0.18	0.82	0.18	0.82	0.18	0.82	0.02	0.29	0.69	0.03	0.31	0.66	0.03	0.30	0.67	0.02	0.32	0.66
hCV2729903	37.14	0.21	0.79	0.21	0.79	0.21	0.79	0.21	0.79	0.05	0.33	0.62	0.04	0.34	0.63	0.04	0.34	0.63	0.04	0.35	0.62
rs2055272	37.29	0.36	0.64	0.36	0.64	0.36	0.64	0.36	0.64	0.11	0.50	0.39	0.13	0.46	0.41	0.14	0.45	0.41	0.12	0.48	0.40
hCV685985	37.30	0.85	0.15	0.85	0.15	0.85	0.15	0.85	0.15	0.71	0.27	0.01	0.73	0.24	0.03	0.72	0.26	0.02	0.74	0.23	0.03
hCV3025711	37.33	0.50	0.50	0.54	0.46	0.53	0.47	0.54	0.46	0.25	0.50	0.25	0.30	0.49	0.22	0.28	0.49	0.22	0.30	0.48	0.22
hCV685941	37.35	0.65	0.35	0.67	0.33	0.66	0.34	0.68	0.32	0.41	0.46	0.12	0.45	0.43	0.12	0.44	0.45	0.12	0.47	0.42	0.11
hCV3025667	37.43	0.68	0.32	0.71	0.29	0.71	0.29	0.71	0.29	0.49	0.39	0.12	0.51	0.39	0.09	0.51	0.40	0.09	0.53	0.37	0.10
rs1757191	37.48	0.49	0.51	0.45	0.55	0.46	0.54	0.44	0.56	0.25	0.48	0.27	0.21	0.48	0.31	0.23	0.47	0.30	0.19	0.50	0.31
rs1739629	38.58	0.37	0.63	0.37	0.63	0.37	0.63	0.37	0.63	0.15	0.43	0.42	0.15	0.45	0.41	0.15	0.45	0.40	0.16	0.42	0.42
rs1332010	38.99	0.30	0.70	0.29	0.71	0.28	0.72	0.28	0.72	0.09	0.43	0.48	0.08	0.42	0.50	0.09	0.38	0.52	0.05	0.47	0.48
rs1004279	39.02	0.74	0.26	0.77	0.23	0.76	0.24	0.79	0.21	0.55	0.38	0.07	0.58	0.38	0.05	0.57	0.38	0.05	0.61	0.35	0.04
rs1541816	39.16	0.48	0.52	0.47	0.53	0.47	0.53	0.48	0.52	0.22	0.52	0.26	0.23	0.48	0.29	0.22	0.49	0.29	0.24	0.48	0.28
rs1014737	41.02	0.11	0.89	0.12	0.88	0.12	0.88	0.11	0.89	0.01	0.20	0.78	0.01	0.22	0.77	0.01	0.23	0.76	0.01	0.19	0.80
hCV2504280	41.05	0.86	0.14	0.85	0.15	0.84	0.16	0.88	0.12	0.74	0.25	0.01	0.72	0.26	0.02	0.71	0.27	0.02	0.76	0.23	0.01
hCV2814735	41.16	0.36	0.64	0.34	0.66	0.34	0.66	0.34	0.66	0.14	0.43	0.43	0.12	0.44	0.43	0.12	0.43	0.44	0.12	0.44	0.44
rs1817538	41.24	0.48	0.52	0.55	0.45	0.55	0.45	0.55	0.45	0.25	0.45	0.30	0.30	0.50	0.20	0.31	0.49	0.20	0.28	0.54	0.18
rs1351835	41.24	0.48	0.52	0.55	0.45	0.55	0.45	0.54	0.46	0.26	0.44	0.30	0.30	0.50	0.21	0.31	0.49	0.20	0.27	0.54	0.19
rs1817537	41.24	0.52	0.48	0.45	0.55	0.45	0.55	0.46	0.54	0.30	0.44	0.26	0.20	0.50	0.30	0.20	0.49	0.31	0.19	0.53	0.27
hCV1620987	41.31	0.27	0.73	0.25	0.75	0.24	0.76	0.26	0.74	0.07	0.39	0.54	0.06	0.38	0.56	0.06	0.36	0.58	0.06	0.40	0.54
hCV3042474	41.32	0.74	0.26	0.76	0.24	0.77	0.23	0.75	0.25	0.55	0.39	0.07	0.57	0.38	0.06	0.59	0.36	0.05	0.55	0.40	0.05
rs742735	41.61	0.78	0.22	0.77	0.23	0.78	0.22	0.76	0.24	0.61	0.34	0.05	0.60	0.35	0.05	0.61	0.34	0.05	0.58	0.36	0.06
rs1410492	41.90	0.75	0.25	0.73	0.27	0.74	0.26	0.69	0.31	0.54	0.41	0.05	0.53	0.40	0.07	0.55	0.39	0.06	0.49	0.41	0.10
rs1160867	42.89	0.53	0.47	0.55	0.45	0.56	0.44	0.54	0.46	0.29	0.49	0.22	0.31	0.48	0.21	0.30	0.53	0.17	0.33	0.41	0.26
rs1105024	43.08	0.05	0.95	0.08	0.92	0.08	0.92	0.07	0.93		0.10	0.90		0.16	0.84		0.17	0.83		0.13	0.87
rs9395	43.15	0.03	0.97	0.06	0.94	0.06	0.94	0.05	0.95		0.06	0.94		0.11	0.89		0.11	0.88		0.09	0.91
rs1111786	43.49	0.98	0.02	0.97	0.03	0.97	0.03	0.96	0.04	0.95	0.05		0.93	0.07		0.94	0.06		0.92	0.08	
rs1418488	44.14	0.34	0.66	0.35	0.65	0.35	0.65	0.35	0.65	0.10	0.46	0.43	0.12	0.47	0.41	0.12	0.47	0.41	0.12	0.46	0.42

For the examination of association of IBD with the SNPs in MHC region 142 SNP markers were genotyped using diallelic discrimination analysis. With this method the genotype of each sample is directly determined so that the genotype and allele frequency in a population can be directly derived. The complete allele and genotype frequencies for the control group, the group of IBD affected individuals and in both disease categories CD and UC are given in Table 8.1. Blanks in one of the genotype fields identify SNPs where only one homozygous and the heterozygous state were observed in the population investigated. In the case population only one affected individual per family was included for the calculation of the frequencies.

8.1.2. Two-point TDT analysis

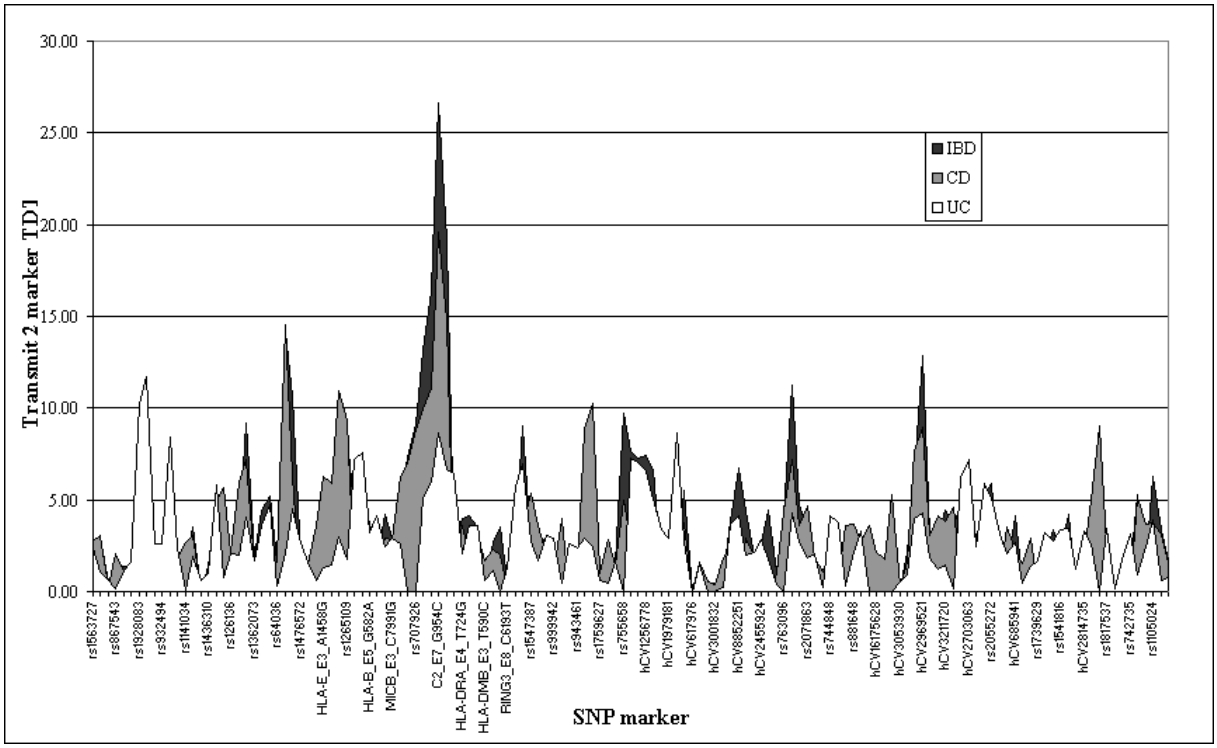

Fig. 8.1: Two point TDT (Transmit) association analysis. The χ^2 was calculated on the basis of the two marker haplotype alleles transmitted from parent to affected offspring compared to the alleles expected from HWE. Significance of a peak at the 95% confidence interval is reached at a χ^2 value of 6.6 with 3 df. Not all markers are shown and the positioning is not according to the physical distance.

From the multipoint TDT analyses only the two-point Transmit results are shown here. The tendency in the results from the Genehunter analysis was very similar. Used for the plot was the global χ^2, which is calculated from all haplotypes of the two marker unit. In the three-marker analysis the situation is reflected, no new association leads could be observed.

8.1.3. Additional linkage analysis results

Linkage analysis using Genehunter was performed for all three disease categories. Here the NPL scores for CD and UC are shown (Fig 8.2 and 8.3). For the analysis of NPL in CD 105 ASPs were available, in the UC analysis only 43. In relation to the number of ASPs in the analysis and that SNPs were the markers employed, which are less informative than microsatellites, an NPL score of 2.5 at the maximum is a good indicator of Linkage to CD in the region. Similar to the NPL in IBD there is a clear decrease towards the centromeric part of the region. To the telomeric side the decrease is less prominent but steeper than in the analysis of IBD.

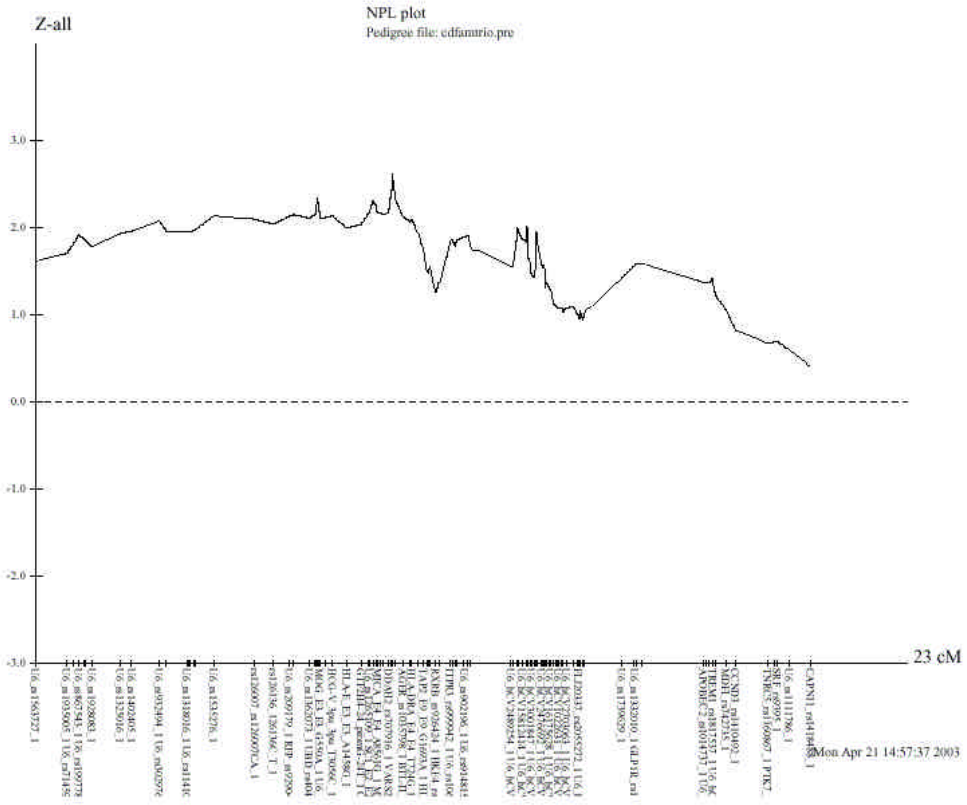

Fig. 8.2: Non-parametric LOD score for CD. The NPL was calculated on the basis of 105 ASPs using the weighed pairs option of Genehunter. All NPL values were highly significant ($p < 0.001$). Z-all identifies the NPL score while on the vertical axis the physically correct positions of the markers are indicated

The NPL score in the UC category is below 1 over the complete region. This is too low to claim linkage to UC in the region. However with the additional 43 ASPs from UC in the IBD analysis the NPL score experienced another increase (compared to the CD results). This indicated the linkage is added though the UC ASPs and is present in UC as well.

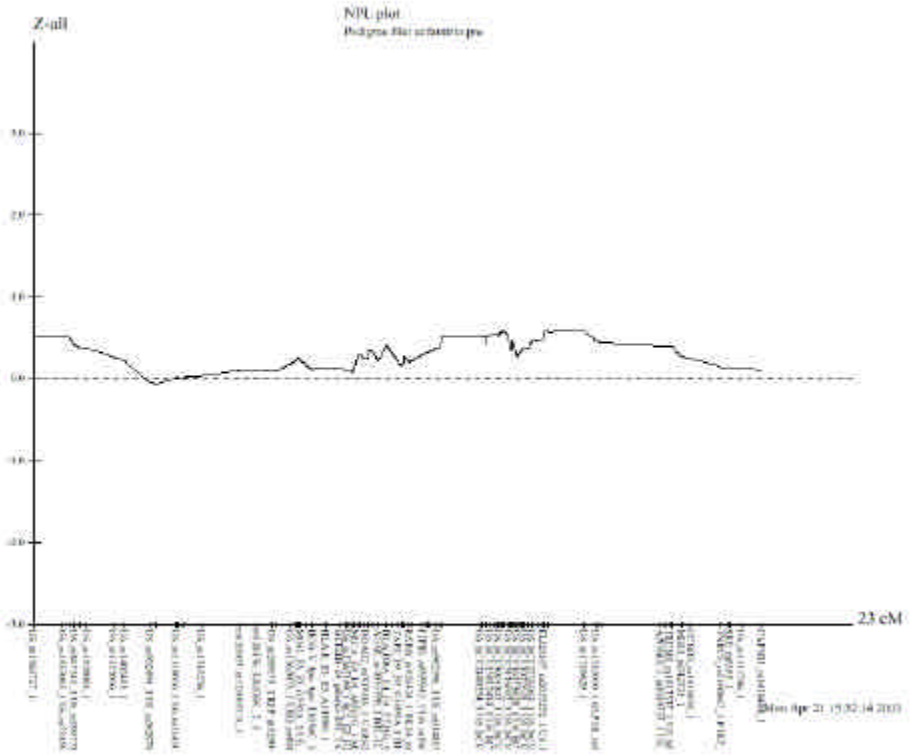

Fig. 8.3: Non-parametric LOD score for UC. The NPL was calculated on the basis of 43 ASPs using the weighed pairs option of Genehunter. Significance was not achieved for the values. For details see Fig. 8.2.

Besides the calculation of the NPL scores the Genehunter program generated a comparison of recombination events expected (theta) from the physical distance and the observed recombinations. Given in the Table 8.2 is the marker and the interval (the marker and the next marker), and the theta expected besides the observed recombination. In most situations the observed recombination was below the expected. The maker intervals where the observed recombination was larger than the expected did not sum up to a significantly higher value. Therefore it can be assumed that the marker order is correct and the used physical distances between markers are close to the real distances.

Table 8.2 Analysis of recombination.

SNP marker name	interval	theta		observed
rs1563727	01-02	0.0096	0.0048	(0.0194/4)
rs1935005	02-03	0.0017	0	(0.0001/4)
rs714591	03-04	0.0016	0	(0.0001/4)
rs867543	04-05	0.0016	0	(0.0001/4)
rs1997781	05-06	0.0003	0	(0.0000/4)
rs1024216	06-07	0.0019	0	(0.0001/4)
rs1928083	07-08	0.0085	0.0001	(0.0004/4)
rs1325016	08-09	0.0034	0.0001	(0.0003/4)
rs1402405	09-10	0.0082	0.0001	(0.0005/4)
rs932494	10-11	0.0021	0	(0.0002/4)
rs302976	11-12	0.0065	0.0001	(0.0005/4)
rs1318016	12-13	0.0002	0	(0.0000/4)
rs1141034	13-14	0.0004	0	(0.0000/4)
rs1747522	14-15	0.001	0	(0.0001/4)
rs1540273	15-16	0.0005	0	(0.0000/4)
rs1436310	16-17	0.0055	0.0001	(0.0004/4)
rs1535276	17-18	0.0122	0.0005	(0.0021/4)
rs126007	18-19	0.0055	0.0003	(0.0011/4)
rs126136	19-20	0.005	0.0003	(0.0010/4)
rs209179	20-21	0.001	0	(0.0002/4)
rs929042	21-22	0.0049	0.0002	(0.0007/4)
rs1362073	22-23	0.0014	0	(0.0002/4)
rs404240	23-24	0.0001	0	(0.0000/4)
rs362536	24-25	0.0001	0	(0.0000/4)
rs64036	25-26	0.0005	0	(0.0001/4)
GABABR1_E7_G141esA	26-27	0.0005	0	(0.0001/4)
MOG_E3_G550A	27-28	0.0004	0	(0.0001/4)
rs1476572	28-29	0.0012	0.0001	(0.0002/4)
rs1569315	29-30	0.0023	0.0001	(0.0004/4)
HCG-V_3pu_T3056C	30-31	0.0043	0.0002	(0.0007/4)
HLA-E_E3_A1458G	31-32	0.0042	0.0001	(0.0003/4)
GTF2H4G-24T	32-33	0.0021	0	(0.0001/4)
CDSN_E1_A66T	33-34	0.0003	0	(0.0000/4)
rs1265109	34-35	0.0001	0	(0.0000/4)
SC1_E2_G2712A	35-36	0.0011	0	(0.0000/4)
HLA-C_E2_G1019A	36-37	0.0008	0	(0.0000/4)
HLA-B_E5_G582A	37-38	0.0004	0	(0.0000/4)
rs2005423	38-39	0.0002	0	(0.0000/4)
MICA_E4_A8561G	39-40	0.0009	0	(0.0001/4)
MICB_E3_C7991G	40-41	0.0008	0	(0.0001/4)
TNFA_G308A_	41-42	0.0015	0	(0.0002/4)
rs707916	42-43	0.0006	0	(0.0001/4)
rs707926	43-44	0.0003	0	(0.0000/4)
HSP70-HOM_G2763A	44-45	0.0002	0	(0.0000/4)
rs539689	45-46	0.001	0	(0.0001/4)
C2_E7_G954C	46-47	0.0021	0.0001	(0.0003/4)
rs1035798	47-48	0.0022	0.0001	(0.0003/4)
BTNL-II_E3_A158G	48-49	0.0004	0	(0.0001/4)
HLA-DRA_E4_T724G	49-50	0.0019	0.0001	(0.0003/4)
HLA-DQA1_E3_C389T	50-51	0.0017	0.0001	(0.0002/4)
TAP2_E9_G1693A	51-52	0.0011	0	(0.0001/4)
HLA-DMB_E3_T590C	52-53	0.0001	0	(0.0000/4)
rs1063478	53-54	0.0004	0	(0.0000/4)
rs206775	54-55	0.0001	0	(0.0000/4)
RING3_E8_C6193T	55-56	0.0002	0	(0.0000/4)
rs2581	56-57	0.0017	0	(0.0001/4)
rs926424	57-58	0.0001	0	(0.0000/4)
rs1547387	58-59	0.0009	0	(0.0000/4)
rs466384	59-60	0.0003	0	(0.0000/4)
TAPBP_E4_G325C	60-61	0.0035	0	(0.0002/4)
rs999942	61-62	0.0004	0	(0.0000/4)
rs10672	62-63	0.0007	0	(0.0000/4)
rs747694	63-64	0.0003	0.0045	(0.0181/4)
rs943461	64-65	0.0023	0.032	(0.1279/4)
rs902196	65-66	0.0012	0.0163	(0.0653/4)
rs914815	66-67	0.0007	0.0103	(0.0411/4)
rs1759627	67-68	0.0123	0.1698	(0.6794/4)
hCV2489254	68-69	0.0002	0.0027	(0.0110/4)
hCV2489242	69-70	0.0011	0.0145	(0.0581/4)
rs755658	70-71	0.0001	0	(0.0000/4)
rs992105	71-72	0.0001	0	(0.0000/4)
rs1320991	72-73	0.0001	0	(0.0000/4)
hCV1256778	73-74	0.0003	0	(0.0000/4)
rs1360780	74-75	0.001	0	(0.0000/4)
hCV15812414	75-76	0.0003	0	(0.0000/4)
hCV1979181	76-77	0.001	0	(0.0000/4)
rs1326752	77-78	0.0003	0	(0.0000/4)
hCV2472420	78-79	0.0003	0	(0.0000/4)
hCV617976	79-80	0.0001	0	(0.0000/4)
hCV617970	80-81	0.0007	0	(0.0000/4)
hCV3001847	81-82	0.0002	0	(0.0000/4)
hCV3001832	82-83	0.0001	0	(0.0000/4)
hCV8852296	83-84	0.0001	0	(0.0000/4)
hCV3001824	84-85	0.0001	0	(0.0000/4)
hCV8852251	85-86	0.0005	0	(0.0000/4)
hCV2455932	86-87	0.0001	0	(0.0000/4)
hCV2455926	87-88	0.0001	0	(0.0000/4)
hCV2455924	88-89	0.0001	0	(0.0000/4)
hCV2455922	89-90	0.0001	0	(0.0000/4)
hCV2455919	90-91	0.0001	0	(0.0000/4)
rs763096	91-92	0.0001	0	(0.0000/4)
rs2007683	92-93	0.0001	0	(0.0000/4)
rs2815812	93-94	0.0001	0	(0.0000/4)
rs2071863	94-95	0.0014	0	(0.0000/4)
hCV2476692	95-96	0.0003	0	(0.0000/4)
hCV2476745	96-97	0.0006	0	(0.0000/4)
rs744848	97-98	0.0001	0	(0.0000/4)
rs763017	98-99	0.0001	0	(0.0000/4)
hCV3247999	99-100	0.0001	0	(0.0000/4)
rs881648	100-101	0.0001	0	(0.0000/4)
rs881647	101-102	0.0001	0	(0.0000/4)
rs1885205	102-103	0.001	0	(0.0000/4)
hCV16175628	103-104	0.0001	0	(0.0000/4)
hCV7516771	104-105	0.0001	0	(0.0000/4)
hCV791696	105-106	0.0002	0	(0.0000/4)
hCV3053930	106-107	0.0007	0	(0.0000/4)
hCV2969554	107-108	0.0009	0	(0.0000/4)
rs1059234	108-109	0.0001	0	(0.0000/4)
hCV2969521	109-110	0.0004	0	(0.0000/4)
hCV1022631	110-111	0.0004	0	(0.0000/4)
hCV1022677	111-112	0.0008	0	(0.0000/4)
hCV3211720	112-113	0.0002	0	(0.0000/4)
hCV2061156	113-114	0.0001	0	(0.0000/4)
hCV2061161	114-115	0.0009	0	(0.0000/4)
hCV2703063	115-116	0.0002	0	(0.0000/4)
hCV2317531	116-117	0.0019	0	(0.0001/4)
hCV2729903	117-118	0.0015	0	(0.0000/4)
rs2055272	118-119	0.0001	0	(0.0000/4)
hCV685985	119-120	0.0003	0	(0.0000/4)
hCV3025711	120-121	0.0002	0	(0.0000/4)
hCV685941	121-122	0.0008	0	(0.0001/4)
hCV3025667	122-123	0.0006	0	(0.0000/4)
rs1757191	123-124	0.0109	0.0002	(0.0008/4)
rs1739629	124-125	0.0041	0.0001	(0.0003/4)
rs1332010	125-126	0.0003	0	(0.0000/4)
rs1004279	126-127	0.0014	0.0001	(0.0002/4)
rs1541816	127-128	0.0186	0.0006	(0.0024/4)
rs1014737	128-129	0.0004	0	(0.0001/4)
hCV2504280	129-130	0.001	0.0001	(0.0002/4)
hCV2814735	130-131	0.0008	0	(0.0002/4)
rs1817538	131-132	0.0001	0	(0.0000/4)
rs1351835	132-133	0.0001	0	(0.0000/4)
rs1817537	133-134	0.0007	0	(0.0002/4)
hCV1620987	134-135	0.0001	0	(0.0000/4)
hCV3042474	135-136	0.003	0.0002	(0.0006/4)
rs742735	136-137	0.0029	0.0002	(0.0006/4)
rs1410492	137-138	0.0099	0.0003	(0.0012/4)
rs1160867	138-139	0.0019	0.0001	(0.0002/4)
rs1105024	139-140	0.0006	0.0003	(0.0014/4)
rs9395	140-141	0.0035	0.0018	(0.0073/4)
rs1111786	141-142	0.0064	0.0034	(0.0135/4)
rs1418488				
total recs = 1.06	real map	= 23.14,	observed	map = 30.49,

8.1.4. Cluster heat-maps of other populations

Here the results of the cluster analysis in the other three populations are shown. While regarded from the physical axis the regions of high LD are similar for all populations of European ancestry. The US-African Americans were less similar and the major high LD areas were smaller but still present. The fragmentation of areas with high LD is greater in this population, which mirrors the otherwise reported shorter haplotypes seen in populations with African ancestry. The differences between the populations from the aspect of clustering are mainly due to a different turning of the branches, while the cluster themselves stay similar.

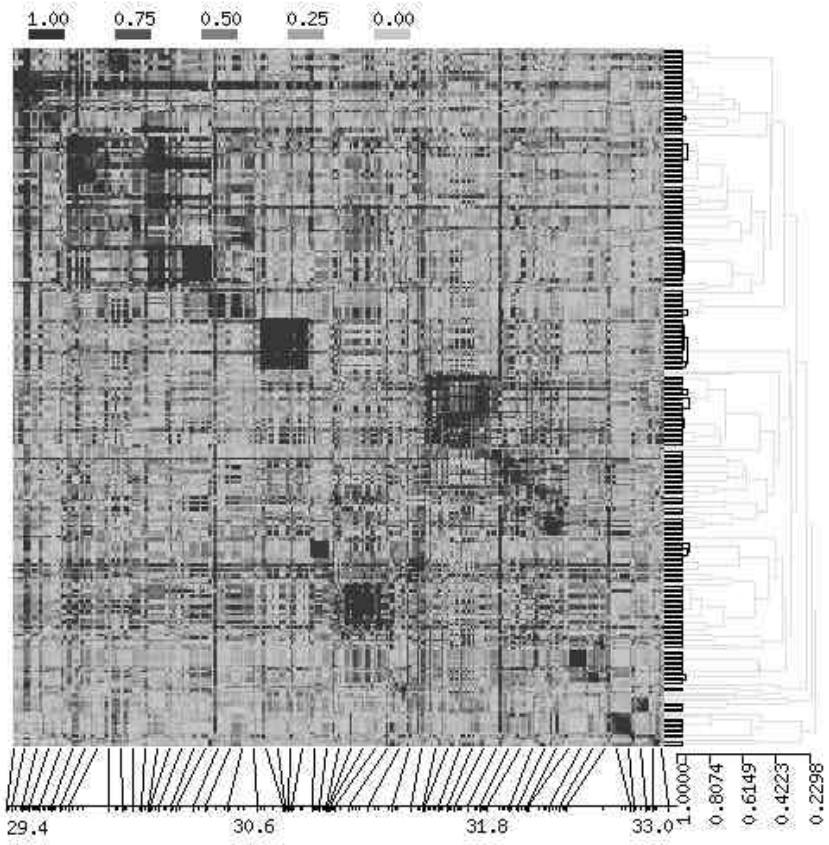

Fig. 8.4: Cluster heat-map of pairwise D' in the UK population.
A total of 320 SNP markers were genotyped in 78 UK individuals. 270 SNPs with minor allele frequencies >5% were analysed. The graph relates the physical map of the region (horizontal axis, with every 5th SNP marked) to the order derived from UPGMA clustering of D' values (vertical axis). The vertical axis details the cluster dendogram with a scale of D' levels shown at the bottom. Levels of pairwise D' are colour-coded from dark grey to light grey (top left corner).

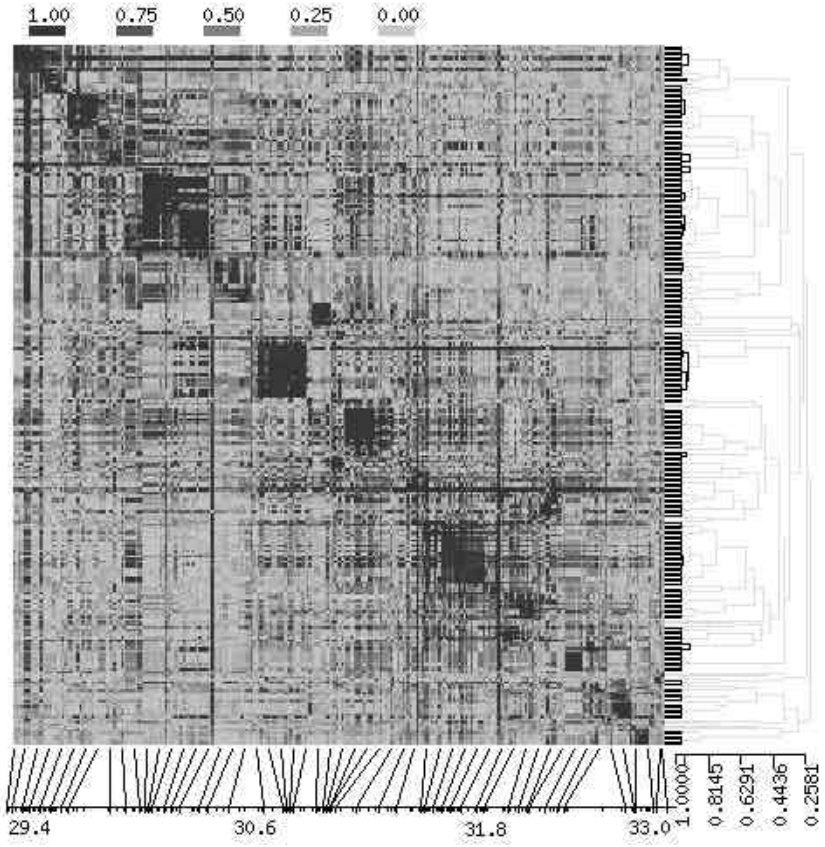

Fig. 8.5: Cluster heat-map of pairwise D' in the Norwegian population.
A total of 320 SNP markers were genotyped in 93 Norwegian individuals. 276 SNPs with minor allele frequencies >5% were analysed. The graph relates the physical map of the region (horizontal axis, with every 5th SNP marked) to the order derived from UPGMA clustering of D' values (vertical axis). The vertical axis details the cluster dendogram with a scale of D' levels shown at the bottom. Levels of pairwise D' are colour-coded from dark grey to light grey (top left corner).

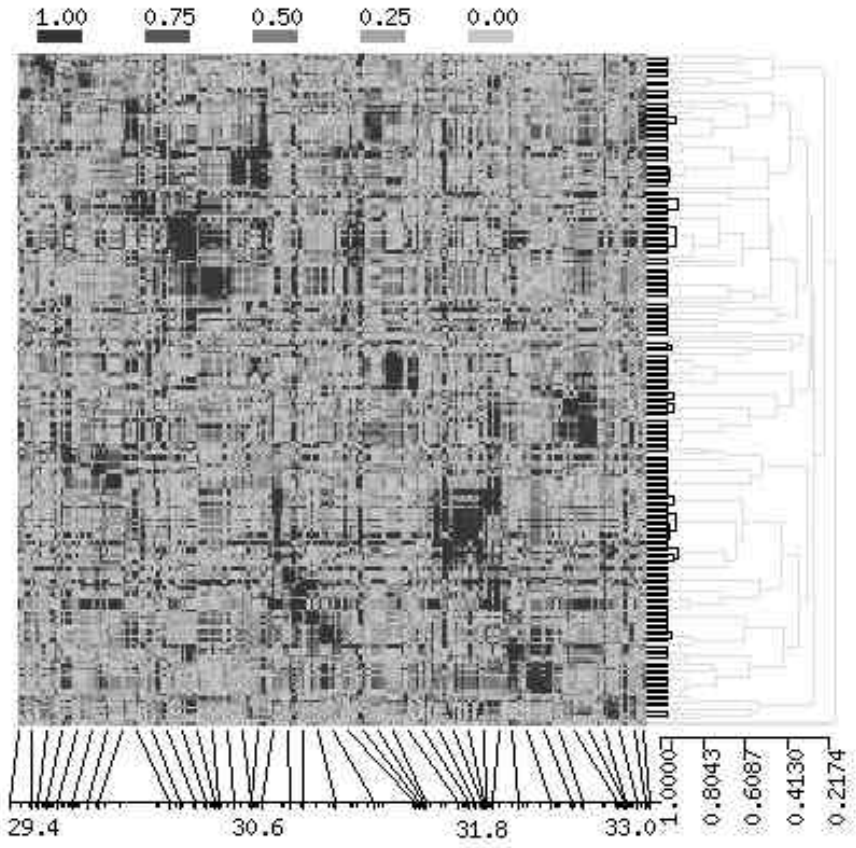

Fig. 8.6: Cluster heat-map of pairwise D' in the US African American population.
A total of 320 SNP markers were genotyped in 45 US African American individuals. 209 SNPs with minor allele frequencies >5% were analysed. The graph relates the physical map of the region (horizontal axis, with every 5th SNP marked) to the order derived from UPGMA clustering of D' values (vertical axis). The vertical axis details the cluster dendogram with a scale of D' levels shown at the bottom. Levels of pairwise D' are colour-coded from dark grey to light grey (top left corner).

8.2. Sequencing primer

Table 8.3a: Primer for PCR amplification and sequencing in MAPK14

primer forward	oligonucleotide	primer reverse	oligonucleotide
MAPK14			
p38_prom_A_F1	TGGTCAGGCTGGTCTCGAACTC	p38_prom_A_R	TCCCGTTCCAGCTGCTGC
p38_prom_B_F1	TGGCTCTTGGAACCGCGA	p38_prom_B_R	GTAGAACGTGGGCCTCTCCTGAG
p38_prom_A_F2	CCCGACCTCAGGTGATCCG		
p38_prom_B_F2	CGACCACTGGAGCCTTAGCG		
p38_prom_A_F3	CTGGGATTACAGGCGTGAGC	p38_prom_A_R2	GCTAAGGCTCCAGTGGTCG
p38_prom_A_F4	AGGCGTGTGCTGCCAGC	p38_prom_A_R3	GCGAGAAGAGAACAATAACTGGAGAC
p38_E1s_F	GAGAGGGTGCGGGTGCAG	p38_E1s_R	CGCTGAGCAGTGGAGCAGG
p38_E1s_F2	GGGCCCACAGGGCCAC		
p38_E1s_F3	GGCGACCAGCGCAAGGT		
p38_E2s_F	AGACCCTTTAATTTGGAAATAGCC	p38_E2s_R	AACGTTTGATGCACTCCAGC
p38_E2s_F2	AAATACCCCAAAATAATATTTAGAACAGC		
p38_E3s_F	CCCGGCCTAGACCCTGACTC	p38_E3s_R	ACCCAAAAAGAGGTGCATGATAGG
p38_E3s_F2	AGATTAAGACCTGATGGGTACTATACTGAG		
p38_E4s_F1	CAGCCTGGCAGTAGAGCGAG	p38_E4s_R	GAACATGGCAGTTTACTGTACTATATGACC
p38_E4s_F2	CTGAAGATAAGAACTTTCTTAGGGGTTCTAG	p38_E4s_R2	AAAGTTGTTTACTTGTTTGTTTAAGCC
p38_E5s_F	CTTACTGATTCATGAAAATGTGCAGC	p38_E5s_R	ATCTTCTTTCTTTAATCAGCTTAAACCTACAG
p38_E5s_F2	TATATGTTATAAATATGACAATAGAAGGTTGG		
p38_E6s_F	TGTAGGGGGAGAATTGATCATGC	p38_E6s_R	ATACGAATTTCAGATAGTCATGGCTCAC
p38_E6s_F2	CATGCATCATAAAGTTGATCCTGTC		
p38_E7s_F	GACCAGGAAATCTCTTTATGAAAGTGC	p38_E7s_R	CCCAAATGCATCCAACATGG
p38_E7s_F2	CTCCTTTTTAATAAGGCAACAGAGG		
p38_E8s_F	GAGTATGATTGTTGAGCCTCAGATGTC	p38_E8s_R	AAGCAAAATACTGAGAGTCTTCCCTAGAG
p38_E8s_F2	TTGAATATTTCACCCCTAGTTACGG		
p38_E9s_F	AGGGATTCACCGTGTTGGCTAG	p38_E9s_R	GATTGGTAGAAGGCAAGTGGCAG
p38_E9s_F2	CATTTTTGTTCTCGGTTGTTCTG		
p38_E10s_F	TGCAGCTGTGGGCTCTCG	p38_E10s_R	TTAAGAACTGTTGGTTATACGGTGGC
p38_E10s_F2	TAGGGGGAATGGAGGGTAGG		
p38_E11s_F	TTAACACTCGTTCCTTAGCAGGACC	p38_E11s_R	TCTCTCCTTTCTCCACTGTACCTGC
p38_E11s_F2	CAAGATTCTTTCTTTGAGAGCAGTAAC		
p38_E12s_F	GTCAAAACTATGTTTGCTCAATAAGGC	p38_E12s_R	ATTCATTCTGCATTAGAGACCTCATCTAC
p38_E12s_R2	TAGCAAGTTGGTTATGACCATTCAG		
p38_E13s_F	GAAGTCAGAGTGCTTGCCAGC	p38_E13s_R	GAAATCTCTGCAACAAGAGGCAC
p38_E13s_F2	CAGCAAAGAGAATAGCCTAAACCTC		

Table 8.3b: Primer for PCR amplification and sequencing in MAPK13

primer forward	oligonucleotide	primer reverse	oligonucleotide
MAPK13			
MAPK13_e1_F1	CAGGTGGGAGCGTAGCAGC	MAPK13_e1_R	AGGTGGCGCTGGCGAG
MAPK13_e1_F2	GAGGAGCGGGCGGC	MAPK13_e1_R2	CTTGAGGTCCGCGGTCAAG
MAPK13_e1_F3	GGAGCGCAGGGCTGGA		
MAPK13_e2_F1	CATATTCTAGGTGGTGGGCCTAGTC	MAPK13_e2_R	ACTGTCCCGAGCTCTCCATACCG
MAPK13_e2_F2	AGCCCAGCTGCCCAGTG		
MAPK13_e3_F1	CTTCCAAGAAGGTTGGAGACATGC	MAPK13_e3_R	CACCCGGAATAGAAACCCTAGACC
MAPK13_e3_F2	GGATGGCCCTGTGCATG		
MAPK13_e4_F1	ACACTGGAGGGTGTCTCTATGACC	MAPK13_e4_R	TGGCCGACCTGTAACAGTGC
MAPK13_e4_F2	CCCTGTTCTGCCCTGAGGTC		
MAPK13_e5_F1	GGGACTTGGAGGCTGGCAG	MAPK13_e5_R	TTGAAGCCTGAAACACTCAGCTCA
MAPK13_e5_F2	GGGACGCCTTGCTGTCTAGAC		
MAPK13_e6_F1	GGGCCCTGAGAGAACACCTAGTC	MAPK13_e6_R	CAAAATCCAGAATCTATGGCACAGG
MAPK13_e6_F2	GGCCTGGCTGAGGTCACAC	MAPK13_e6_R2	GGATAAATAGAAAAGGAACTTGCTGAG
MAPK13_e7_F1	TGAGGACTGTGAACTGAAGGTGAGTG	MAPK13_e7_R	GAGGGAGAGTGTCACCTGGACTG
MAPK13_e7_F2	CCCAGAGGCGGGATAGGC	MAPK13_e7_R2	CACAGACCAGATGTCCACTGCAG
MAPK13_e7_F3	GAGTGGGCTGCAGGCTCAG		
MAPK13_e8_F1	GTGATCCTCAGCTGGATGCACTAC	MAPK13_e8_R	TCTACAGGGAGTATGTATCACTCTCACA
MAPK13_e8_F2	CAGACAGGTCAGTGGTCAATGC		
MAPK13_e9_F1	AGAGCAGGAGGTGGTTAGGCAG	MAPK13_e9_R	TGTGGGAAGGTGCTGGTGAG
MAPK13_e9_F2	GGAGCCTGGATGATCAGTTGC	MAPK13_e9_R2	GATTAAGACAAGAGTGCCTGCCAG
MAPK13_e9_F3	TCAGCCTGGCACGTTGGA		
MAPK13_e10_F1	TTGTGGAAGAAGGCTCACCAGC	MAPK13_e10_R	AGCAGGTCCGCAGCTGCA
MAPK13_e10_F2	GGCACTCTTGTCTTAATCTCACCG		
MAPK13_e11_F1	CCCCAGGGTGAGTCTCAGAGC	MAPK13_e11_R	TAGGGCTGGCAGAGGCCAG
MAPK13_e11_F2	GCCTCTCAGCGTATCCCAGAG		
MAPK13_e12_F1	TTGAGCTGACTTCGCTCTCCATG	MAPK13_e12_R1	GCAATCGCACCAGCTGCAG
MAPK13_e12_F2	GCTCTCCATGCTGTCTCTGAAG	MAPK13_e12_R2	ACAGAGCACAAGGCCTCAGC
MAPK13_e12_F3	CCAGCCCTACCTGCCACCTC	MAPK13_e12_R3	CGCAGTCTCCGAGCCTGC

Table 8.3c: Primer for PCR amplification and sequencing in TREM1 and BRPF3

primer forward	oligonucleotide	primer reverse	oligonucleotide
TREM1			
TREM1_prom_F1	TGTGGGCCTGACTCTCTTCACTAC	TREM1_prom_R	CTCATCCTTCTGTGCACCAGC
TREM1_prom_F2	ACACTAAACTGGATGTGAATGGAG		
TREM1_e1_F1	GGCCTCATATCCTGTTGTGCA	TREM1_e1_R	CCTGTGCTCTGAAGCTTCAACAG
TREM1_e1_F2	CAACTTTCCGAAGCCTCTAGGTC		
TREM1_e2_F1	TGGAGGCCTCAAGAACCTCATC	TREM1_e2_R	TGCTGATGCTACCACCACTGC
TREM1_e2_F2	TGAAATATGGGTGGTTGGACAAG		
TREM1_e3_F1	CCTACCCGTCCATCAACTCATC	TREM1_e3_R	CCTCACTTTCACATCCTCTCAGC
TREM1_e3_F2	TTTCATCCATACATCCATACATCCA		
TREM1_e4_F1	GGACAGGGACCTATACTCTCCACTG	TREM1_e4_R	GTGGAATGAAGGTCCAAGCATG
TREM1_e4_F2	TTTTGGCAATGAGACAGACTGA		
BRPF3			
BRF3-e0F1	GGGAGAAGACCGCGCTCC	BRF3-e1R1	TTCTGCCGGGACTTCCGAC
BRF3-e1F1	GAGGAAGCCTCGTCGGAAGTC	BRF3-e1R2	GTATCTGTGGGACAGCAAGCATG
BRF3-e1F2	GAAGCATCAGTGAGACTGGCGAT	BRF3-e2R1	GGGACTGCAGATGGGAGTGC
BRF3-e2F1	ACAAGATCTGTAGTGGTCTCTCCTTTCAG	BRF3-e3R1	TCTTCCGAATCAGCTCAATCAGC
BRF3-e3F1	GAGAGCAGGATGAGAAGACAAGTGC	BRF3-e4R1	CACTCAAGTTGACTGGTTCTGCGA
BRF3-e4F1	AAGTCCAGCAGGCTGCCATG	BRF3-e5R1	AAGTCTTCCAATTTGGGTGACTCG
BRF3-e5F1	ATTACCTGGAATTCATATCCAAGCCA	BRF3-e6R1	TGCTCCAGCACAGCTGCATC
BRF3-e6F1	TTGTCCCCAGAGGTGCAGCT	BRF3-e7R1	TCCCCTTCGCTCTCACTACAGC
BRF3-e7F1	AACTGCCTCCTCCGCCAAC	BRF3-e8R1	CTGCAAGCTTCAAATGCCAGTC
BRF3-e8F1	CATGACCAACGGCTTTGGAA	BRF3-e9R1	GAGCCATTGTAGTCAGAGTCTCCGT
BRF3-e9F1	ACGCAGCCGTGGGAAGC	BRF3-e10R1	CAAGGCAGGGTAGGAGGGGTAG
BRF3-e10F1	CCTTTGAAGACCGCGGAGAC	BRF3-e11R1	CCAGGTGCGCTTGTTGTCA
BRF3-e11F1	GGGAGGGCCTCCTGCAC	BRF3-e12R1	TCAGGTGGATCATCGCACG
BRF3-e11F2	GCTTCCAAGGGACAAAGTCCTG	BRF3-e12R2	CCTGCCAGGAGAACCATGC

8.3. Blood collection centres

The following organisations, medical centres and local physicians participated in the contacting of affected individuals and their families and the unrelated control individuals and the collection of blood samples and questionnaires:

The cooperation of the German Crohn's and colitis foundation (DCCV e. V.); Prof. Raedler / Hamburg, Prof. Kruis, Köln; Dr. Theuer, Heilbronn; Dr. Meckler, Gedern; Prof. Lochs, Dr. Wedel, T. Herrmann, Berlin; Dr. Herchenbach, Recklinghausen; Prof. Scheurlen, Würzburg; Dr. Demharter, Augsburg; Dr. Simon, Munich; Dr. Purrmann, Moers; Dr. Jessen, Kiel; Dr. Zehnter, Dortmund; Dr. Lübke, Dr. Weismüller, Koblenz; Dr. Eiche, Denkendorf; Dr. Schönfelder, Aache; Prof. Fleig, Halle, all in Germany.

Dr. Wewalka, Linz; Dr. Knofloch, Wels, both in Austria.

Prof. Dr. Ulrich R. Fölsch, Prof. Dr. Stefan Schreiber, Dr. Susanna Nikolaus, Dr. Tanja Kühbacher, 1st Department of General Internal Medicine at the University Clinic Schleswig-Holstein, Campus Kiel (Kiel, Germany);

Dr. Peter Nürnberg, Charité University Hospital (Berlin, Germany);

Prof. Dr. Neppert, Transfusion Medicine at the University Clinic Schleswig-Holstein, Campus Kiel (Kiel, Germany).

Prof. Dr. Christopher Mathew, Dr. Alastair Forbes, King's College School of Medicine, Guy's Hospital, and the ST. Mark's Hospital London (UK).

Prof. Dr. SJH van Deventer, Dr. Pieter CF StokkersAcademic Medical Center, Amsterdam (The Netherlands).

Prof. Dr. Trevor Winter and Marc Kidd, Groote Schuur Hospital, University of Cape Town, Department of Internal Medicine (Cape Town, South-Africa).

Prof. Dr. Won Ho Kim, Yonsei University, College of Medicine (Seoul, South-Korea).

Prof. Dr. Morten Vatn; Rikshospitalet Oslo, Norway,

8.4. Questionnaire

Patient

Fragebogen zur Familienstudie bei chronisch entzündlichen Darmerkrankungen

Persönliche Daten

männlich❑	weiblich❑	Geburtsjahr:

In verschiedenen Volksgruppen treten die chronisch entzündlichen Darmerkrankungen in verschiedener Häufigkeit und Ausprägung auf

Ethnische Herkunft (z.B. deutsch, türkisch, italienisch, polnisch):			
Woran sind Sie erkrankt?		nicht erkrankt	❑
Morbus Crohn	❑	Crohn/Colitis nicht entschieden	❑
Colitis ulcerosa	❑	Collagene colitis	❑
Welche Ihrer Verwandten hat ebenfalls Morbus Crohn oder Colitis Ulcerosa?			
Welcher Ihrer Verwandten hat Symptome wie Sie selbst auch, ist aber bisher nicht untersucht worden?			
Wie viele Geschwister haben Sie? (Bitte mit Alter und Geschlecht angeben, wenn verstorben bitte Jahr und Alter angeben)		_____ _____ _____	
Wie viele Kinder haben Sie selbst (Bitte mit Alter und Geschlecht angeben)		_____ _____ _____	
Rauchen/Rauchten Sie Ja ❑ Nein❑		Wenn ja, wie viele Jahre: von.....bis..... Wie viele Zigaretten pro Tag:	

In einer englischen Studie wurde ein Zusammenhang zwischen dem Stadtleben und dem Risiko an Morbus Crohn zu erkranken hergestellt:

Wo sind Sie aufgewachsen ?		Wo lebe Sie heute?	
Dorf	❑	Dorf	❑
Kleinstadt	❑	Kleinstadt	❑
Großstadt	❑	Großstadt	❑
In einem Haushalt mit wie vielen Personen wuchsen Sie auf?_____ In einem Haushalt mit wie vielen Personen leben Sie heute?			
Welcher Sanitärkomfort herrschte in Ihrer Kindheit?			
Fließendes Wasser	Ja ❑	Nein ❑	
Fließendes Warmwasser	Ja ❑	Nein ❑	
Wasserklosett	Ja ❑	Nein ❑	
Zentralheizung	Ja ❑	Nein ❑	

1

Wurden Sie gegen Tuberkulose (Tbc) geimpft?	Ja❑	Nein ❑
Wurden Sie gegen Masern geimpft?	Ja❑	Nein ❑
Sind Sie als Kind an Masern erkrankt	Ja❑	Nein ❑
Besteht bei Ihnen Asthma? wenn ja, seit wann:	Ja❑	Nein ❑
Besteht bei Ihnen eine Neurodermitis/Ekzem?	Ja❑	Nein ❑
Besteht bei Ihnen eine rheumatische Organerkrankung? Wenn ja, welche: Seit wann:	Ja❑	Nein❑

Die folgenden Fragen beziehen sich nur auf erkrankte Familienmitglieder: Ihre individuelle genetische Veranlagung kann möglicherweise den Krankheitsverlauf beeinflussen:

Welche Darmteile sind bei Ihnen betroffen?			
Dünndarm	❑	Übergang vom Dünndarm zum Dickdarm❑	
Teile des Dickdarms	❑welche:		
ganzer Dickdarm:	❑		
Enddarm	❑		
genauere Angaben:			
Haben Sie während Ihrer Erkrankung Fisteln entwickelt?	Ja❑	Nein❑	
Wenn ja, welche:			
Enddarm (Analfisteln)	Ja❑	Nein❑	weiß nicht❑
Fisteln zwischen Darmschlingen	Ja❑	Nein❑	weiß nicht❑
Fisteln vom Darm zur Haut	Ja❑	Nein❑	weiß nicht❑
Fisteln zu anderen Organen	Ja❑	Nein❑	weiß nicht❑
Haben sich bei Ihnen im verlauf der Erkrankung Stenosen (Darmverengungen) ausgebildet?	Ja ❑	Nein ❑	
Haben/Hatten Sie extraintestinale Beschwerden?	Ja ❑	Nein ❑	
Gelenkschmerzen während der Schübe	Ja ❑	Nein ❑	
Hautentzündungen	Ja ❑	Nein ❑	
Augenentzündungen	Ja ❑	Nein ❑	
Gallengangsentzündungen	Ja ❑	Nein ❑	
Seit Welchem Lebensjahr ist die Diagnose klar?			
Seit welchem Lebensjahr hatten Sie erste Beschwerden?			
Wie wurde Ihre Erkrankung bestätigt?			
Feingewebeschnitt	Ja ❑	Nein ❑	
Darmspiegelung	Ja ❑	Nein ❑	
Darmröntgen	Ja ❑	Nein ❑	
Wie viele - und welche – Operationen sind aufgrund Ihrer Erkrankung notwendig gewesen?			

Haben/Hatten Sie einen künstlichen Darmausgang? Ja ❑ Nein ❑

Wie viele stationäre Krankenhausaufenthalte sind bei Ihnen notwendig gewesen?

		Wie lange insgesamt im Krankenhaus?	
Nie	❑		
Weniger als alle 2 Jahre	❑	Weniger als 1 Monat	❑
ca. 1x pro Jahr	❑	1-3 Monate	❑
1-3 x pro Jahr	❑	3-6 Monate	❑
Mehr als 3 x pro Jahr	❑	Mehr als 6 Monate	❑

Wie lange haben Sie insgesamt Kortison > 10mg pro Tag nehmen müssen?

Mehr als ein Jahr	❑
bis zu einem Jahr	❑
3-6 Monate	❑
1-3 Monate	❑
Weniger als 1 Monat	❑
Noch nie	❑

Wie viele Schübe haben Sie in etwa pro Jahr?

Weniger als alle 2 Jahre	❑
ca. 1x pro Jahr	❑
1-3 x pro Jahr	❑
Mehr als 3 x pro Jahr	❑

Wann war der letzte Schub:

Sind Sie wegen Ihrer chronisch-entzündlichen Darmerkrankung vorzeitig berentet worden? Ja ❑ Nein ❑

Wenn ja, in welchem Lebensalter?

Was würden Sie in der Betreuung von Patienten mit chronisch-entzündlichen Darmerkrankungen verändern?

Nach Eingang der Proben wird Ihre Adresse und Ihr Name aus unseren Unterlagen gelöscht. Aufgrund vieler Rückfragen von Patienten möchten wir Sie hier aber noch einmal fragen.

Bitte Löschen Sie meine Adresse und Namen nach Eingang sofort:
Ja ❑ Nein ❑

Ich möchte weiter über die allgemeinen Ergebnisse der Studie (die in ca. 3 Jahren zu erwarten sind) auf dem Laufenden gehalten werden:
Ja ❑ Nein ❑

Vielen Dank für Ihre Mitwirkung!
Bei Rückfragen melden Sie sich bitte gerne unter Telefon 0431 597 1373

This questionnaire was completed by all affected individuals from the European and the South African populations but as well by the unaffected family members, each adapted to the mother tongue. All information was entered into the database. The severity markers used in the analysis were taken from the information provided by the individual in the questionnaire

8.5. Index of figures and tables

Tables **Page**

9. Curriculum vitae

PERSONAL INFORMATION

Name:	Annette Stenzel
Date of birth:	29th July 1971
Place of birth:	Bad Wildungen, Germany
Citizenship:	German
Marital status:	Married

EDUCATION

May 1992	General qualification for university entrance (Abitur); Hans-Thoma-Gymnasium, Lörrach, Germany
October 1993	1. Intermediate diploma in Biology I (botany, chemistry, mathematics), University of Basel, Switzerland
October 1994	2. Intermediate diploma in Biology I (zoology, organic chemistry, biochemistry, physics), University of Basel, Switzerland
September 1997	Diploma in Biology I, (medical parasitology-epidemiology, population-biology, biology of vertebrates, plant-medicine); University of Basel, Switzerland
since March 1998	Ph.D. student in the Mucosal Immunology Group, Dept. of General Internal Medicine, University Clinic Schleswig-Holstein, Campus Kiel, Germany

RESEARCH EXPERIENCE

March 1994 - July 1995	Drosophilalaboratories, Zoological Institute, University of Basel, Switzerland (Dr. Ranka Junge-Berberovi´c) assistance: *Life History consequences of Adaptation to Fluctuating Temperatures in Drosophila*.
July 1995 - October 1995	Swiss Tropical Institute (STI), Basel, Switzerland / Ifakara Health Research and Development Centre, Ifakara, Tanzania (Prof. Dr. Marcel Tanner): *Ecology of Biomphalaria pfeifferi, the intermediate host of Schistosoma mansoni*.
November 1996 - September 1997	Swiss Tropical Institute (STI), Basel, Switzerland / Ifakara Health Research and Development Centre, Ifakara, Tanzania (Prof. Dr. Marcel Tanner) Diploma thesis: "*Operational Research on the Control of human Schistosomiasis: ecological and epidemiological aspects*"
September 1997 - February 1998	Swiss Tropical Institute (STI), Basel, Switzerland, (Dr. Marianne Ostermeyer): *Development of a routine-diagnostic tests (ELISA) to recognize infection with Strongyloides stercoralis*.
since March 1998	Mucosal Immunology Group, Dept. of General Internal Medicine, University Clinic Schleswig-Holstein, Campus Kiel, Germany (Prof. Dr. Stefan Schreiber): *Development of an SNP map in the peri-MHC Region on the human Chromosome 6 as a tool to identify candidate genes for inflammatory bowel disease*

10. Declaration (Erklärung) and publication list

Apart from the advice of my supervisors, this thesis is wholly the result of my own work. No part of it has been submitted to any other board for another qualification. Most of the results have already been published (see below).

Hiermit erkläre ich, dass diese Dissertation – abgesehen von der Beratung durch meine akademischen Lehrer – nach Inhalt und Form meine eigene Arbeit ist. Sie hat weder im Ganzen noch zum Teil an anderer Stelle im Rahmen eines Promotionsverfahrens vorgelegen. Ein Teil der Ergebnisse dieser Arbeit wurde bereits veröffentlicht (siehe unten).

Kiel, ... (Annette Stenzel)

PUBLICATIONS RELATED TO THIS THESIS

PAPER

Hampe J, Shaw SH, Saiz R, Leysens N, **Lantermann A**, Mascheretti S, Lynch NJ, MacPherson AJ, Bridger S, van Deventer S, Stokkers P, Morin P, Mirza MM, Forbes A, Lennard-Jones JE, Mathew CG, Curran ME, Schreiber S. (1999) Linkage of inflammatory bowel disease to human chromosome 6p. *Am J Hum Genet*. Dec;65(6):1647-55.

Lantermann A, Hampe J, Kim WH, Winter TA, Kidd M, Nagy M, Folsch UR, Schreiber S. (2002) Investigation of HLA-DPA1 genotypes as predictors of inflammatory bowel disease in the German, South African, and South Korean populations. *Int J Colorectal Dis*. Jul;17(4):238-44.

CONGRESS ABSTRACTS

Lantermann A, Winter TA, Kim WH, Nagy M, Kidd M, Köpke B, Hampe J, Fölsch UR, Schreiber S (2000)
Investigation of HLA- DPA1 association with IBD in European, South-African and South Korean populations
Digestive Disease Week (DDW), Orlando, FL, USA; May 2000
- Poster - *Gastroenterology* 118 (4, Suppl. 2): A334.

Lantermann A, Gong J, Hampe J, Schreiber S (2001)
Development of a Single Nucleotide Polymorphism Map in the HLA-Region on Chromosome 6
German Human Genome Project (DHGP), Jahrestagung, Braunschweig Germany, November 2001
- Poster -

Lantermann A, Wjst M, Jenisch S, Gong J, Stoll M, Hampe J Schreiber S (2002)
LD in the HLA Region on Chromosome 6 using a set of 120 SNP markers
National Genome Research Network (NGFN) Symposium: Genomics of chronic inflammatory disorders Kiel, Germany, July 2002
- Poster -

Lantermann A, Lu T, Dong P, De La Vega FM, Gong J, Goll P, Günther SM, Hampe J, Stoll M, Krawczak M, Schreiber S (2002)
Association mapping of the HLA region in inflammatory bowel disease
Symposium of the National Genome Research Network (NGFN) and the German Human Genome Project (DHGP), Berlin, Germany November 2002
- Oral presentation -

Stenzel A, Lu T, Koch WA, Hampe J, De La Vega FM, Su X, Dong P, Guenther SM, Krawczak M, Schreiber S.
The Landscape of Linkage Disequilibrium in the human MHC Region
European Human Genetics Conference (ESHG), Birmingham, UK, May 2003
- Poster - *Europ. J. Hum. Genet.* (11, Suppl. 1): P743

11. Acknowledgements

First of all I would like to express my sincerest thanks to Prof. Dr. Hans-Dieter Flad, who supported and advised me in spite of his retirement and subsequent move to South Germany, and his continuing full agenda.

I deeply appreciate the expertise and advice I received from Prof. Dr. Stefan Schreiber who has given me the opportunity to work under excellent conditions and with access to large resources of patient material and laboratory, equipment fundamental for the success of this work. His conception of an applied genetic epidemiology and his continuous encouragement and incentive has opened many prospects to me. My thanks go as well to Dr. Jochen Hampe who has put a lot of effort in the structure of the laboratory platform, the development of the database and the collection of patients, and whose scientific work provided the basis of this thesis.

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.

This work would not have been possible without the willingness of people from all over the world to donate blood and genetic material for scientific work, and the medical workers who helped with their clinical expertise and time to collect the material and the medical information relevant for the analyses. I am grateful for the help and commitment I have experienced.

The expert assistance from the laboratory technicians is most appreciated. Tanja Engler and Illona Urbach deserve special thanks for the continuous high quality work of preparing DNA samples for the use in the high throughput format. The work of Tanja Wesse, Birthe Petersen and Hebke Hinz in the genotyping of many SNPs, with an expert eye for the quality was outstanding, as were their readiness to help and their humour when time was scarce. Tam Ho Kim, Marlies Zornow and Nadine Tepe are thanked for their qualified support in the sequencing process.

I thank Marcus Will for his never ending work to keep the server-based computer network running and his helpfull comments on a thousand questions on various software problems. He and the other members of the IT team made it possible to administrate and analyse the large amount of data.

The advice from Prof. Dr. Michael Krawczak and the assistance in the statistical analysis of linkage disequilibrium by Tim Lu is gratefully acknowledged. I appreaciate very much the advice I received through Dr. Monika Stoll.

I wish to thank Francisco De La Vega for the statistical analysis of LDU and the fruitful cooperation with Applied Biosystems (Foster City, CA, USA), and as well Applied Biosystems Deutschland (Überlingen, Germany) where Simone Guenther was an outstanding collaboration partner.

My colleagues at the Mucosal Immunology Group deserve my sincerest thanks for the discussions concerning biological and technical aspects on the bumpy way to high-throughput. They also provided me with constant refreshment of the English language and generated a pleasant research atmosphere. Besides his sense for humour Georg Wätzig showed me how to finish a Ph.D. in really short time.

I am deeply indebted to Birte Köpke, who from being a colleague in the first place, became a friend with an always open ear and heart, who endured many bad moods. Since the day we first met she has taken major influence in my life in more than one way.

I am very grateful to my parents who supported me in all my intentions with time, love and trust, and provided me with the foundation to dare and to enjoy life. During the time of my thesis I came to know and love my parents-in-law who welcomed me to their family with open arms and provided me with a second home.

My final thanks go to my husband Hauke Stenzel who is more to me and did more for me than words can express.