

FINNISH RED CROSS BLOOD TRANSFUSION SERVICE AND  
FACULTY OF SCIENCE, DEPARTMENT OF BIOSCIENCES,  
DIVISION OF GENETICS, UNIVERSITY OF HELSINKI, FINLAND

**GENETIC SUSCEPTIBILITY TO CELIAC DISEASE:  
HLA-UNLINKED CANDIDATE GENES**

**PÄIVI HOLOPAINEN**

ACADEMIC DISSERTATION

To be publicly discussed, with the permission of the Faculty of Science of the University of Helsinki, in the Nevanlinna Auditorium of the Finnish Red Cross Blood Transfusion Service, Kivihaantie 7, 00310 Helsinki, on May 24<sup>th</sup> 2002, at 12 o'clock noon.

Helsinki 2002

**ACADEMIC DISSERTATIONS FROM THE  
FINNISH RED CROSS BLOOD TRANSFUSION SERVICE  
NUMBER 45**

**Supervised by**

Docent Jukka Partanen, PhD  
Department of Tissue Typing  
Finnish Red Cross Blood Transfusion Service  
Helsinki, Finland

**Reviewers**

Docent Erkki Savilahti, MD, PhD  
Hospital for Children and Adolescents  
University of Helsinki  
Helsinki, Finland

Docent Pentti Tienari, MD, PhD  
Department of Neurosciences  
University of Helsinki  
Helsinki, Finland

**Opponent**

Professor Kimmo Kontula, MD, PhD  
Department of Medicine  
University of Helsinki  
Helsinki, Finland

ISBN 951-97576-9-4 (Print)  
ISBN 952-5457-00-1 (PDF)  
ISSN 1236-0341  
<http://ethesis.helsinki.fi>  
Tummavuoren Kirjapaino Oy  
Vantaa 2002

Ihminen on tullut viisaaksi silloin,  
kun hän tajuaa,  
ettei tiedä mistään mitään.  
*-tuntematon*



*To my children-to-be*



# CONTENTS

LIST OF ORIGINAL PUBLICATIONS .....	8
ABBREVIATIONS .....	9
ABSTRACT .....	10
REVIEW OF THE LITERATURE .....	12
<b>CELIAC DISEASE – GLUTEN SENSITIVITY</b> .....	12
<i>CLINICAL ASPECTS</i> .....	12
<i>DIAGNOSTICS</i> .....	14
<i>EPIDEMIOLOGY</i> .....	14
<i>PATHOGENESIS</i> .....	15
<b>GENETICS OF CELIAC DISEASE</b> .....	17
<i>MAJOR HISTOCOMPATIBILITY COMPLEX AND HLA-MOLECULES</i> .....	17
<i>HLA-ASSOCIATED DISEASES</i> .....	19
<i>HLA-LINKED GENES IN CELIAC DISEASE</i> .....	19
<i>HLA-UNLINKED GENES IN CELIAC DISEASE</i> .....	21
<b>GENETIC ANALYSES OF COMPLEX GENETIC DISEASES</b> .....	21
<i>GENETIC CHALLENGES OF MULTIFACTORIAL DISEASES</i> .....	22
<i>DEFINING THE GENOTYPES</i> .....	23
<i>GENOME-WIDE SCREENINGS AND CANDIDATE GENE APPROACHES</i> .....	24
<i>GENETIC LINKAGE ANALYSES</i> .....	24
<i>GENETIC ASSOCIATION ANALYSES</i> .....	26
<i>STATISTICAL SIGNIFICANCE</i> .....	27
<b>GENOMEWIDE SCREENINGS AND CANDIDATE GENE APPROACHES IN CELIAC DISEASE</b> .....	28
<i>GENOME-WIDE ANALYSES</i> .....	28
<i>CANDIDATE GENES</i> .....	29
<i>CANDIDATE GENES CD28, CTLA4 AND ICOS ON CHROMOSOME 2q33</i> .....	29
AIMS OF THE STUDY .....	32
MATERIALS AND METHODS .....	33
STUDY ETHICS .....	33
STUDY SUBJECTS .....	33
GENETIC MARKERS .....	34
HLA TYPING .....	34
DATA ANALYSIS .....	34
RESULTS .....	36
LINKAGE IN CANDIDATE REGIONS 2q33, 5q, 11q AND 15q26 (I-III) .....	36
ASSOCIATION AND TDT RESULTS ON CANDIDATE GENES .....	36
LINKAGE HETEROGENEITY (I) .....	37
LINKAGE DISEQUILIBRIUM WITHIN THE CTLA4 GENE (IV) .....	39
GENOME-WIDE LINKAGE STUDY (V) .....	40
SUSCEPTIBILITY TO CELIAC DISEASE DUE TO HLA FACTORS (I, V) .....	42
DISCUSSION .....	44
CANDIDATE GENES IN CELIAC DISEASE .....	44
THE ROLE OF HLA - DOSE-EFFECT AND SEX-MODULATED RISK DUE TO DQ2 .....	50
FAILURES IN REPLICATING LINKAGE - REASONS? .....	51
CELIAC DISEASE AS A COMPLEX GENETIC DISORDER .....	55
CONCLUDING REMARKS .....	56
ACKNOWLEDGEMENTS .....	58
REFERENCES .....	60
ORIGINAL PUBLICATIONS	

## LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, which are referred to in the text by Roman numerals (I-V).

- I **Holopainen P**, Mustalahti K, Uimari P, Collin P, Mäki M and Partanen J. Candidate gene regions and genetic heterogeneity in gluten sensitivity. *Gut* 48:696-701, 2001.
- II Susi M\*, **Holopainen P\***, Mustalahti K, Mäki M and Partanen J. Candidate gene region 15q26 and genetic susceptibility to coeliac disease in Finnish families. *Scand J Gastroenterol* 36, 372-374, 2001.
- III **Holopainen P**, Arvas M, Sistonen P, Mustalahti K, Collin P, Mäki M and Partanen J. CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to celiac disease. A linkage and family-based association study. *Tissue Antigens* 53, 470-475, 1999.
- IV **Holopainen PM** and Partanen JA. Technical note: linkage disequilibrium and disease-associated CTLA4 gene polymorphisms. *J Immunol* 167:2457-2458, 2001.
- V Liu J, Juo SH, **Holopainen P**, Terwilliger J, Tong X, Grunn A, Brito M, Green P, Mustalahti K, Mäki M, Gilliam TC and Partanen J. Genomewide linkage analysis of celiac disease in Finnish families. *Am J Hum Genet* 70:51-59, 2002.

The articles in this thesis have been reproduced with the permission of the copyright holders.

\*Authors with an equal contribution on the paper.

**ABBREVIATIONS**

Ag	antigen, antigenic peptide
AGA	anti-gliadin antibodies
APC	antigen presenting cell
ARA	anti-reticulin antibodies
bp	base pair
CD	celiac disease
cM	centiMorgan
CTLA4	cytotoxic T-lymphocyte associated antigen 4
DH	dermatitis herpetiformis
EMA	anti-endomysium antibodies
ESPGAN	European Society for Paediatric Gastroenterology and Nutrition
FH	familial hypercholesterolemia
GSE	gluten-sensitive enteropathy
HLA	human leucocyte antigen
IBD	identical by descent
IBS	identical by state
ICOS	inducible co-stimulator
IDDM3-10	susceptibility loci in type I (insulin-dependent) diabetes mellitus
IL	interleukin
kb	kilobase, 10 <sup>3</sup> base pairs
LD	linkage disequilibrium
LDL	low density lipoprotein
LOD	logarithm of odds
Mb	megabase, 10 <sup>6</sup> base pairs
MHC	major histocompatibility complex
MLS	maximum likelihood score
NPL	non-parametric linkage
PCR	polymerase chain reaction
RFLP	restriction fragment length polymorphism
SNP	single nucleotide polymorphism
TcR	T lymphocyte receptor
TDT	transmission/disequilibrium test
Th1, Th2	T-helper lymphocyte type 1, 2
tTG	tissue transglutaminase

## ABSTRACT

Celiac disease (CD) is a relatively common autoimmune type of inflammation of the small intestine also manifesting with various extraintestinal symptoms. The disease is triggered by dietary gluten of wheat, barley and rye, but the pathogenetic mechanisms are not completely resolved. The only established genetic risk factors so far are the human leucocyte antigens (HLA) DQ2 and DQ8, either of which is carried by nearly all patients. HLA-linked genes can not, however, explain the whole genetic susceptibility for this multifactorial disease. Several attempts to localize the HLA-unlinked risk factors have included both genome-wide linkage studies, as well as candidate gene approaches on positional or functional candidate genes, and resulted in partly inconsistent results between the populations.

In this thesis we searched for the HLA-unlinked susceptibility genes of celiac disease. In study I, two positional candidate regions 5q and 11q suggested in a previous Italian genome scan were investigated in 102 Finnish families with affected sib-pairs. Evidence for linkage in these regions was supported. The candidate region 15q26, reported to be linked with CD in populations of the British Isles, was also tested (II), but no linkage in this region was observed. Among the most interesting functional candidate gene loci for CD is chromosome 2q33, harboring genes for the cytotoxic T-lymphocyte associated antigen 4 (CTLA4), CD28 and inducible co-stimulator (ICOS). These molecules regulate the T-lymphocyte activation processes and have been suggested to play a role in the pathogenesis of several autoimmune diseases. Significant linkage of 2q33 with CD was observed, but no allelic association with the CTLA4 gene was found (III). A very high level of linkage disequilibrium between the widely studied CTLA4 polymorphisms was found (IV), which is important to note in the search of the functional risk allele primarily associated with the disease under study. The genome-wide linkage analysis (V) on 60 Finnish families with affected sib-pairs suggested six HLA-unlinked susceptibility loci; 1p36, 4p15, 5q31, 7q21, 9p21-23 and 16q12. The regions 4p, 5q and 7q were further studied with additional markers and 38 additional CD families, but the linkage in these regions was not markedly improved in the larger set of families. Although showing only suggestive linkage, these regions located close to those reported previously at least in some independent linkage studies. The strongest evidence from multiple studies has been accumulated to chromosome 5q which harbors functionally relevant candidate genes encoding Th2 type cytokines.

Consistent with other genome-wide scans in CD, the strongly linked HLA region at 6p21.3 appeared to be the major risk locus even in the Finnish population in which a founder effect and lower level of heterogeneity can be assumed. Dose effect of the carried HLA risk alleles was also supported in our family sample. Linkage heterogeneity in the candidate loci due to

the disease phenotype and sex was suggested, and females carrying HLA-DQ2 in these families had an increased risk of CD compared to males. This could implicate a possibly stronger role of HLA-unlinked risk factors in families with affected males.

The relatively weak linkage evidence and the observed discrepancies between studies and populations in the linked loci points either to a higher rate of false positive results, or a higher level of heterogeneity or other confounding factors which are known to complicate the genetic studies in multifactorial diseases. They also indicate that the HLA-unlinked loci are likely to play only a minor and possibly a much more complex role in the CD susceptibility, which further affects the power of available mapping methods. The figure is typical to most of the other autoimmune diseases showing multifactorial etiology, and the problems and future tasks for the genetic studies of these diseases are discussed.

## REVIEW OF THE LITERATURE

### CELIAC DISEASE – GLUTEN SENSITIVITY

Celiac disease (CD, also known as gluten sensitive enteropathy, GSE, OMIM 212750) can be defined as a permanent intolerance to dietary gluten, resulting in an autoimmune type of injury in the small intestinal mucosa, and sometimes in other tissues, in genetically susceptible individuals. The developing villus atrophy often leads to malabsorption of the nutrients and various clinical symptoms. Thus far, the only cure is a strict, life-long gluten-free diet. The pathogenesis of CD is partly unclear, but the immune system is assumed to play an important role in it. Susceptibility to CD is inheritable, showing features of a complex or multifactorial genetic disease. Confounding factors like genetic and phenotypic heterogeneity, epistasis and gene-environment interactions typical to these diseases challenge the mapping of the genes involved in them. The only susceptibility genes for CD established so far encode HLA-DQ molecules, crucial players during immune activation. The search for other genes that would explain the remaining genetic component of CD has recently started.

### CLINICAL ASPECTS

Celiac disease was first described in 1888 by Samuel Gee who reported on chronic malabsorption of ingested food and described many of the classical symptoms of CD (Gee, 1888). In infancy typical symptoms are chronic diarrhea, steatorrhea, abdominal distension and failure to thrive (Schmitz, 1992), and in adult patients diarrhea, weakness, malaise and weight loss (Howdle and Losowsky, 1992). Over the last few decades, however, the gastrointestinal symptoms have become rarer, the clinical picture has been altered to milder and atypical forms, and the age at diagnosis has increased (Mäki et al., 1988; Collin et al., 1999). Diagnosis, which is based on small intestinal biopsy, disease-specific serum antibodies and the clinical picture, can be made at any age. Several non-abdominal symptoms are common, among which iron deficiency, short stature, delayed puberty, osteoporosis and dental enamel defects may at least partly result from the malabsorption of nutrients. Furthermore, infertility and miscarriages in women, liver diseases and neurological complications can be found (Mäki and Collin, 1997). The reason for the relatively quick change in the clinical picture must obviously be environmental – an effect of a longer period of breast feeding or timing of gluten introduction to infant diet have been suggested (Auricchio et al., 1983; Ivarsson et al., 2000). Other explanations are naturally the diagnostic improvement and the consequent recognition of the milder and atypical forms of the disease.

Celiac disease can also manifest in the skin as dermatitis herpetiformis (DH), an itching and blistering rash which responds to gluten-free diet (Fry et al., 1973). Most DH patients have also CD specific changes in their small bowel mucosa (Marks et al., 1966; Reunala et al., 1984). Silent CD is an asymptomatic form of the disease, which is mainly diagnosed by screening among e.g. first degree relatives of CD patients or other riskgroups. These patients have only mild or no clinical symptoms at the time of diagnosis, but disease specific autoantibodies and various degree of villous atrophy are usually found (Ferguson et al., 1993). Both DH and silent CD can occur in families having patients with classical CD, and DH-CD monozygous twin pairs have been described (Hervonen et al., 2000). This can indicate that the exact disease manifestations are not solely dependent on genetic factors. However, this feature is difficult to investigate, because once the gluten free diet has started it will effectively prevent the onset of other forms of the disease.

Many autoimmune diseases co-occur with celiac disease. CD patients have an increased risk to develop type I diabetes, Sjögren syndrome, autoimmune thyroid diseases and juvenile rheumatoid arthritis (Cooper et al., 1978; Collin et al., 1994) and conversely, the risk of CD is higher among patients suffering from these diseases (Collin et al., 1989&1994; Lepore et al., 1996; Cronin et al., 1997; Sategna-Guidetti et al., 1998; Iltanen et al., 1999). At least tenfold risk of CD have been observed among individuals with selective IgA deficiency compared to population in general, and *vice versa* (Savilahti et al., 1971; Savilahti et al., 1985; Collin et al., 1992; Cataldo et al., 1998). Interestingly, recent screening studies have confirmed the increased risk of CD among patients with Down's (Bonamico et al., 2001), William's (Giannotti et al., 2001) and Turner syndromes (Ivarsson et al., 1999), which makes the chromosomes 21, 7q and X interesting as harboring potential candidate genes for CD, although the CD association with these rare diseases can also be due to the severe metabolic unbalance and disturbed immunity typical to these diseases.

The above mentioned associations between the autoimmune diseases and CD may be at least partly due to the shared HLA susceptibility alleles associated with many of them. Other shared genetic risk factors, such as the CTLA4 gene or its homologies on chromosome 2q33 may have a role in several autoimmune diseases (Kristiansen et al., 2000), and clustering of genomic regions showing genetic linkage to autoimmune diseases has been observed (Becker et al., 1998). Non-genetic factors like the prolonged pathophysiological immune activation state may also decrease the overall threshold for autoimmunity. The risk of developing other autoimmune diseases has been claimed to be higher in CD patients diagnosed in adulthood (Ventura et al., 1999), which could indicate that long gluten exposure itself could predispose to other autoimmune diseases. However, a recent study did not fully support this finding (Sategna-Guidetti et al., 2001). Further studies on the role of gluten exposure in these diseases are needed.

## **DIAGNOSTICS**

Wheat was identified as the dietary trigger of CD in 1950 by Willem-Karel Dicke (Dicke, 1950), and the later studies also showed the toxicity of rye and barley (Anand et al., 1978). The storage protein fraction of cereals contains gluten proteins, and the ethanol-soluble prolamins of gluten are the triggering agents of CD. These prolamins are called gliadin in wheat, hordein in barley and secalin in rye. In the late 1950's the diagnosis of CD was also improved by the introduction of a method for small intestinal biopsies (Shiner, 1957). The European Society for Paediatric Gastroenterology and Nutrition (ESPGAN) provided criteria for CD diagnosis in 1970 which included an observed villus atrophy during normal gluten containing diet, and the healing of the small bowel mucosa during gluten-free diet (Meeuwisse, 1970). The original requirement of a third biopsy after gluten-rechallenge is not supported nowadays, but the small bowel biopsy is still the gold standard of the diagnosis. In developing celiac disease the first histological change in the small intestinal mucosa is the infiltration of the epithelium and subsequently the lamina propria by lymphocytes. This is followed by the hyperplasia of the crypts, with partial, subtotal and finally total villous atrophy, leading to the flat mucosa typical to CD (Marsh, 1992).

The diagnosis of celiac disease is also supported by specific serum antibody tests. These tests are useful also to monitor the strict adherence to a gluten-free diet, as well as in screening and in the search for silent CD patients among first degree relatives and other risk groups (Troncone and Ferguson, 1991; Mäki, 1995; Dieterich et al., 2000). IgA- and IgG-class anti-gliadin-antibodies (AGA) were the first diagnostic antibodies recognized and are still widely used, although their sensitivity and specificity to celiac disease is low (Stern, 2000). Instead, the novel tests for autoantibodies against extracellular matrix components show higher sensitivity and specificity. Methods for detecting IgA-class anti-reticulin- (ARA) and anti-endomysium- (EMA) antibodies are available, together with tests for autoantibodies against tissue transglutaminase (tTG) which is now known to be the major autoantigen for the endomysial antibodies (Dieterich et al., 1997).

## **EPIDEMIOLOGY**

Celiac disease is a relatively common disease in European populations, affecting all age groups. It is twice as common in females than in males (Logan, 1992), also a typical feature of many other autoimmune diseases. The prevalence in Europe was earlier estimated to be 1/1000, although regional differences were found (Greco et al., 1992). Several recent screening studies revealing the previously undiagnosed patients with silent CD, have however indicated a higher prevalence of even 1/100 (Catassi et al., 1994; Johnston et al., 1997; Kolho et al., 1998; Meloni et al., 1999; Korponay-Szabo et al., 1999). The inability to pick up the milder

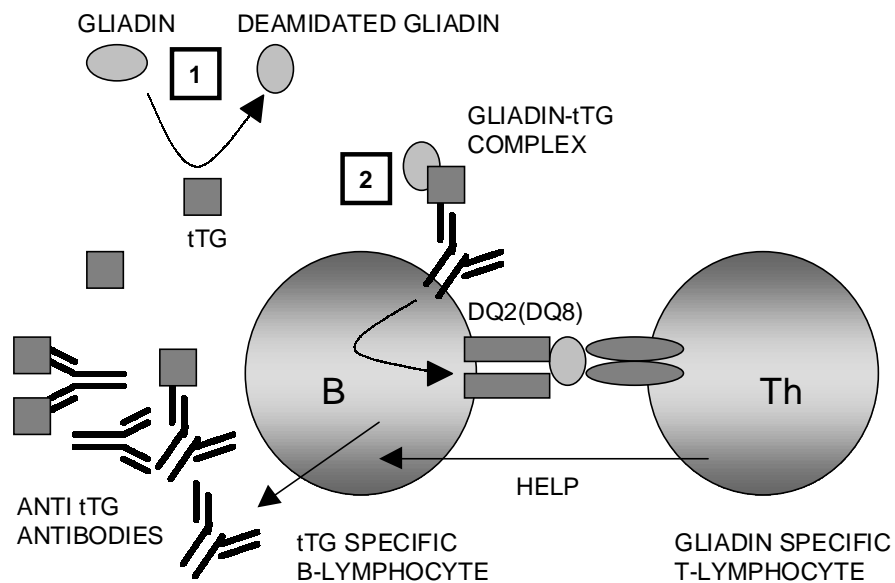
forms probably explains why CD was earlier reported to be rare in USA, although the prevalence of DH was similar to Europe (Smith et al., 1992). Recent screening studies have indeed shown prevalence of 1/250 in USA (Not et al., 1998). Environmental factors such as early exposure to gluten, the amount of gluten intake and the duration of breast feeding may induce fluctuation in the incidence of clinical CD among infants (Ivarsson et al., 2000). Whether the overall prevalence of CD among the population is affected by these factors when asymptomatic patients among all age groups are included is still an open question.

## **PATHOGENESIS**

Pathogenesis of the mucosal lesion in celiac intestine is still partly unknown, although strong evidence points to the T-lymphocyte mediated immune mechanisms, with activation signs of both cellular and humoral responses (Schuppan, 2000). The strong, almost absolute association of CD with HLA-DQ2 and DQ8 suggests that the disease-causing antigen is presented by these molecules (Sollid and Thorsby, 1993). In the intestinal biopsies from CD patients, but not those from controls, CD4+ T-lymphocyte clones reactive to gliadin can be found. Furthermore, these T-lymphocytes recognize the gliadin peptides presented by DQ2 or DQ8, but not by other HLA alleles (Lundin et al., 1993&1994; Molberg et al., 1997). The proinflammatory Th1 type of cytokines secreted by these activated T-lymphocytes could then drive the mucosal damage either directly or by indirect effects on upregulation of HLA expression on enterocytes, activation of cytotoxic T-lymphocytes, increased Fas expression on epithelial cells leading to their apoptosis (Maiuri et al., 2001) or secretion of matrix degrading enzymes by fibroblasts (Pender et al., 1997). The number of intraepithelial lymphocytes bearing the  $\gamma\delta$ -T-lymphocyte receptor is increased in both CD and DH patients (Halstensen et al., 1989; Savilahti et al., 1992), but their role in the pathogenesis is unknown.

Several DQ2 or DQ8 specific epitopes among gliadin peptides have been found (Sjöström et al., 1998; van de Wal et al., 1998b; Anderson et al., 2000; Arentz-Hansen et al., 2000). Interestingly, the T-lymphocyte recognition of most of them depends on an enzymatic modification by tissue transglutaminase (tTG). tTG plays a multifunctional role in the stabilization of the extracellular matrix and in tissue repair, and its extracellular activity is increased during inflammation and in the celiac mucosa (Molberg et al., 2000). tTG appears to play a dual role in celiac disease (Figure 1). It strongly enhances the antigenicity of the gliadin peptides by deamidation of glutamine residues into glutamic acid (Molberg et al., 1998; van de Wal et al., 1998a). tTG is also the major autoantigen against which the highly specific autoantibodies observed in the sera of untreated CD patients are directed (Dieterich et al., 1997). tTG can form cross-links between proteins by covalent bonds between lysin and glutamine residues and the tTG-gliadin complexes have been proposed to direct the autoimmune antibody formation. According to one hypothesis (Sollid et al., 1997), these complexes could be taken up by B-

lymphocytes specific for tTG, resulting in the processing and presentation of both tTG and gliadin peptides by these cells. The help from gliadin-specific T-lymphocytes, which accumulate in celiac intestinal mucosa, could then drive the anti-tTG antibody secretion by these B-lymphocytes. The role of anti-tTG antibodies in the pathogenesis of CD is unclear. Preliminary evidence exists that they could block the essential function of tTG in the proteolytic activation of transforming growth factor (TGF)- $\beta$ . This could then disturb the epithelial cell differentiation which is dependent on TGF- $\beta$  (Halttunen and Mäki, 1999).



**Figure 1.** Dual role of tissue transglutaminase (tTG) in celiac disease. tTG can both (1) deamidate specific gliadin peptides enhancing their binding affinity to DQ2 molecules, and (2) form complexes between itself and gliadin. These complexes could drive the autoantibody production by tTG specific autoreactive B-lymphocytes, with the help from gliadin specific, DQ2 (or DQ8) restricted T-helper lymphocytes.

Due to the strong HLA-association and the formation of autoantibodies, celiac disease presents many features of an autoimmune disease. Compared to many other disorders of this type, CD can indeed serve as a relatively simple model disease for them. The autoantigen (tTG), the highly specific autoantibodies against it and the functional role of the associated HLA-molecules in the disease pathogenesis have been characterized. Importantly, the major environmental trigger, gluten, is also known and encountered practically by all individuals. However, celiac disease can not be strictly defined as a classical autoimmune disease since the loss of immunologic tolerance to the self-antigen is secondary and dependent on gluten.

## **GENETICS OF CELIAC DISEASE**

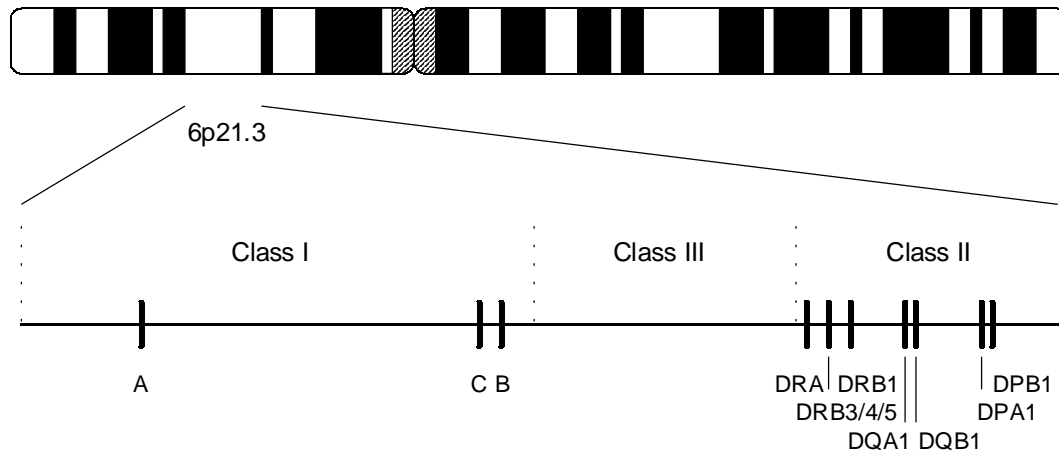
Celiac disease is a typical multifactorial disease in which both several genetic and environmental factors are needed for the disease onset. The familial component is evident, the disease prevalence among the first degree relatives is 10% (Ellis, 1981; Mäki et al., 1991), giving the relative risk of 10, when assuming the prevalence of 1/100 in the general population. Speaking in favor of genuine genetic susceptibility is the fact that the concordance among monozygous twins has been reported to be over 70% (Polanco et al., 1981; Bardella et al., 2000) and two recent studies reported an even higher figure (Hervonen et al., 2000; Greco et al., 2002). Among dizygotic twins the concordance is only 11% which is at the same level as the risk for siblings (Greco et al., 2002). The concordance among monozygous twins appears one of the highest in the field of complex diseases - and it might even be higher if the subjects were followed for a longer time, since the age at onset can vary in terms of decades between twins. The only definitive genetic risk locus thus far is human leucocyte antigen (HLA) DQ at the major histocompatibility complex (MHC) on chromosome 6p21.3.

### ***MAJOR HISTOCOMPATIBILITY COMPLEX AND HLA-MOLECULES***

The main function of the adaptive immune system is to effectively recognize foreign antigens and to eliminate pathogens, at the same time still maintaining tolerance for self antigens and normal flora. The capability of the immune system to recognize a huge number of antigenic peptides, processed and presented by antigen presenting cells (APC), is based on the variability in highly polymorphic HLA genes encoding molecules with slightly different antigen binding specificities. These HLA-antigen complexes are recognized by T-lymphocyte receptors, which are also extremely polymorphic.

HLA genes are among the most polymorphic genes known in man, and they all are located within a gene cluster called the major histocompatibility complex (MHC) spanning 4 megabases (Mb) on chromosome 6p21.3 (Figure 2). The MHC region harbors over 200 genes (The MHC sequencing consortium, 1999) of which many, but not all are related to the function of the immune system. The region can be divided into three classes: the telomeric part carries the class I HLA genes A, B and C, and the centromeric part the genes encoding the class II HLA-antigens DR, DQ and DP (Trowsdale, 1996). The class III MHC region harbors a diverse collection of genes, many of which are involved in the immune system, e.g. the cluster of genes encoding many components of the serum complement system (Aguado et al., 1996).

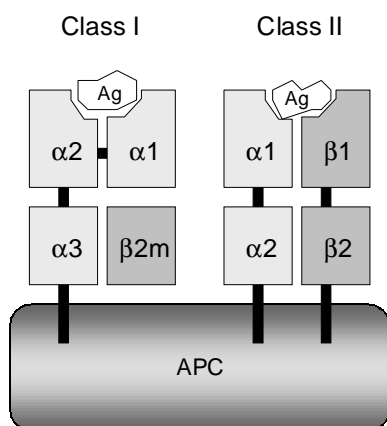
Both class I and II HLA-antigens are transmembrane molecules characterized by an extracellular peptide binding groove (Figure 3). The amino acid variability between the HLA-alleles is concentrated in the residues constructing this groove, which is relevant in respect to the



**Figure 2.** Location of the HLA genes within the MHC gene cluster on chromosome 6p21.3.

ability of the HLA molecules to bind a variety of antigenic peptides. The class I HLA-molecules HLA-A, -B and -C are heterodimers composed of one transmembrane  $\alpha$ -chain, which is encoded by HLA locus and features three extracellular domains and a non-polymorphic  $\beta$ 2-microglobulin ( $\beta$ 2m) encoded by a non-HLA locus. Domains  $\alpha$ 1 and  $\alpha$ 2 of the  $\alpha$ -chain form the groove, which binds antigenic peptides, typically 8-10 amino acids in length. Class II HLA molecules DR, DQ and DP are heterodimers of one  $\alpha$ - and one  $\beta$ -chain,  $\alpha$ 1 and  $\beta$ 1 domains forming the peptide binding groove. Both the  $\alpha$ - and  $\beta$ -chains are encoded by HLA-linked genes. The class II molecules bind peptides which are up to 20 amino acids in length.

The class I HLA-molecules are expressed by nearly all cell types. By these molecules, cells present intracellularly synthesized peptides to cytotoxic CD8+ T-lymphocytes, therefore monitoring the pathogenic changes due to viral infections or transformation of the cell. Instead,



**Figure 3.** Structure of HLA class I and II molecules. The extracellular domains forming the antigenic peptide (Ag) binding groove are illustrated. APC, antigen presenting cell.

the expression of class II HLA molecules is mainly restricted to professional antigen presenting cells of the immune system, such as macrophages, monocytes, B-lymphocytes and dendritic cells. Class II molecules bind extracellular antigens taken in and processed by these cells and present the HLA-antigen complex to CD4+ T-helper lymphocytes. The T-lymphocytes then guide the cell-mediated (Th1 type) or humoral (Th2 type) immune response against the extracellular pathogens, such as bacteria or dietary components.

The multiple HLA genes with highly variable alleles form a broad range of alternative HLA genotypes, or tissue types, of an individual. Along with the additional variation due to *trans* combinations of class II heterodimers in heterozygotes, the HLA system defines the range of antigens against which the immune reactions can be targeted. On the other hand, HLA molecules have a key role in tolerance induction to the body's self antigens, breakage of which could cause a pathogenic autoimmune attack against self proteins and tissues. The repertoire of mature T-lymphocytes, which present a huge variety of T-lymphocyte receptors (TcRs) for effective recognition of pathogens, is defined in the thymus: the potentially autoreactive T-lymphocytes, which react exceptionally strongly with HLA-self peptide complexes presented by thymocytes are eliminated during their maturation.

### **HLA-ASSOCIATED DISEASES**

Certain HLA alleles are found to be associated with diseases, either in a protective or predisposing manner (Hall and Bowness, 1996). Most of these diseases are autoimmune disorders. Several hypotheses have been made to explain these associations. First, the distinct binding properties of HLA molecules could cause differences in the strength of peripheral immune response to some critical antigen. Second, these binding differences between the alleles could affect the tolerance induction to a certain self peptide in the thymus. Third, molecular mimicry between the pathogenic and self peptides may cause an autoimmune reaction triggered by infection. Fourth, the pathogen may also use certain HLA molecule as a receptor, although no convincing evidence for this is available. Finally, the association may also result from the strong linkage disequilibrium (LD) between the associated HLA allele and a closely located primary risk gene. This is the case for example in hereditary hemochromatosis (Feder et al., 1996) or congenital adrenal hyperplasia (White et al., 1984). An exceptionally high level of LD characterizes the MHC region, i.e. certain alleles exist in the same haplotypes far more frequently than expected by their allelic frequencies.

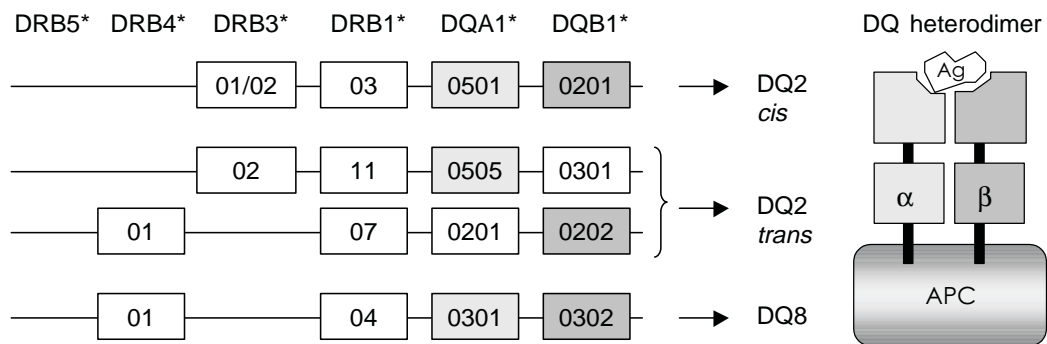
### **HLA-LINKED GENES IN CELIAC DISEASE**

Strong linkage disequilibrium can complicate the fine-mapping of disease associated genes. HLA association of celiac disease was confirmed in 1972, when a significant excess of alleles

REVIEW OF THE LITERATURE

HLA-A1 and HLA-B8 were found in patients, compared to healthy controls (Falchuc et al., 1972; Stokes et al., 1972). Soon after, an even stronger association was found for class II region allele Dw3 (presently DRB1\*03), which often coexists with HLA-B8 (Keuning et al., 1976; Ek et al., 1978). Later, alleles DR7 (DRB1\*07) and DR5 (DRB1\*11) were also found to be associated (DeMarchi et al., 1979; Mearin et al., 1983; Trabace et al., 1998). Today, the primary associated locus has been proven to be HLA-DQ; this is based both on functional and genetic evidence (Sollid, 2000).

HLA-DQ2 molecule associates with CD in all studied populations, being one of the strongest HLA associations among autoimmune type diseases. In most populations, over 90% of CD patients carry the DQ2 heterodimer encoded by alleles DQA1\*05 and DQB1\*02 (Sollid and Thorsby, 1993). DQ2 can be formed *in cis*, i.e. both DQA1 and DQB1 risk alleles are located in the same chromosome (Figure 4). The DQ2 *cis* haplotype is more frequent among CD patients of Northern Europe, and this haplotype invariably carries also the DRB1\*03 allele. In Southern Europe the DQ2 molecule is also frequently formed *in trans* in individuals heterozygous for haplotypes DRB1\*11-DQA1\*05-DQB1\*03 and DRB1\*07-DQA1\*02-DQB1\*02. Most of the CD patients negative for DQ2 carry either the DRB1\*04-DQB1\*0302 (DQ8) haplotype or they are positive for only one of the alleles coding for DQ2 (Spurkland et al., 1992; Polvi et al., 1998).



**Figure 4.** DR-DQ haplotypes carrying the DQA1 and DQB1 alleles which encode the  $\alpha$  and  $\beta$  chains, respectively, of the heterodimers DQ2 and DQ8 associated with celiac disease.

Although one copy of DQ2 is sufficient for disease onset, a dose dependence of DQ2 molecule has been observed (Ploski et al., 1993; Congia et al., 1994). Interestingly, individuals heterozygous to DRB1\*03 and DRB1\*07 haplotypes have been claimed to be at a higher risk of CD than the DRB1\*03-DQ2 homozygotes, suggesting that in addition to two copies of DQB1\*02 present in both genotypes, an additional risk factor in the DRB1\*07 haplotype might be present (Meddeb-Garnaoui et al., 1995; Fernández-Arquero et al., 1995). One suggested

candidate (Clot et al., 1999b) was the DRB4 gene (DR53), which is also present in the DRB1\*04-DQ8 haplotype (Figure 4), but the genetic data do not completely support the DRB4 hypothesis (Partanen, 2000). Some DPB1 alleles have also showed preliminary association to CD (Polvi et al., 1996), and haplotype transmission tests from DQB1\*02 homozygous parents to affected offspring have suggested different transmission patterns, suggesting that additional DQ2 linked risk factors might exist (Polvi et al., 1997; Lie et al., 1999; Karell et al., 2002). Even if these non-DQ risk factors in the HLA region existed, the DQ2 molecule still has the major functional role in pathogenesis. Also, the concordance studies among twins and HLA identical sibs clearly demonstrate that HLA unlinked risk factors are needed for disease onset.

### **HLA-UNLINKED GENES IN CELIAC DISEASE**

Although the alleles HLA-DQ2 or -DQ8 are needed for the onset of celiac disease, they cannot explain the whole genetic susceptibility of the disease. First, they are both common in healthy individuals showing carrier frequencies of 22% and 24% in the Finnish population, respectively (Partanen and Westman, 1997). Second, over 70% concordance among monozygous twins is one of the highest among complex diseases (Polanco et al., 1981; Hervonen et al., 2000; Bardella et al., 2000; Greco et al., 2002), pointing to strong genetic impact and leaving only minor space for variation in environmental risk factor(s) other than gluten. However, in siblings sharing identical HLA haplotypes the CD concordance is only around 30%, suggesting importance of genes located outside HLA in disease susceptibility (Mearin et al., 1983; Petronzelli et al., 1997). The contribution of HLA factors to the entire familial risk to develop CD has been estimated to be only 1/3, which includes both genetic and environmental factors shared by family members (Rotter and Landaw, 1984; Risch, 1987; Petronzelli et al., 1997; Bevan et al., 1999).

### **GENETIC ANALYSES OF COMPLEX GENETIC DISEASES**

All genetic linkage and association analyses basically test the correlation of a given phenotype with the given genotype. Since this correlation is very strong or even complete for the diseases caused by a single gene and showing simple dominant or recessive inheritance pattern, the identification of the genetic cause of these Mendelian disorders is nowadays a relatively straight-forward task. The gene affected usually carries a number of mutations, each resulting in a defective function of protein and causing the characteristic phenotype of the disease. Instead, the defining of the genetic basis of common, multifactorial diseases is much more complicated (Lander and Schork, 1994; Risch, 2000).

## **GENETIC CHALLENGES OF MULTIFACTORIAL DISEASES**

Several factors complicate the identification of the individual susceptibility genes of complex traits. First, the diseases are typically multifactorial, i.e. several environmental and genetic factors are involved in the disease onset. Second, epistatic interaction between these susceptibility genes as well as gene-environment interactions are likely to exist. Third, heterogeneity for the genetic factors occur between, and even within, populations. This heterogeneity can be either allelic, i.e. the disease phenotype is influenced by different allelic variants of the same gene, or locus heterogeneity, i.e. susceptibility alleles of alternative genes can result in the same phenotype. Due to these factors, it is likely that the genetic factors determine only the general susceptibility to the disease, but can not directly cause it. This means that different genetic backgrounds can result in the same phenotypic trait. Therefore, the specific risk attributable to a single gene may be dependent on, e.g. environmental factors, or allelic variation in other genes.

These underlying confounding factors in multifactorial diseases affect the power of available statistical methods for genetic studies. The inheritance model can not be accurately determined from the segregation of the disease in families. Also, the correlation between the phenotype and the given genotype at a single locus is likely to be low due to two phenomena: the penetrance of a single susceptibility allele is nearly always incomplete, meaning that only part of the carriers are affected; and on the other hand, phenocopies are common, i.e. individuals affected without the susceptibility allele. Furthermore, the frequency of this allele in a population can not be *a priori* estimated by the frequency of the disease, which is possible for monogenic traits.

In addition to this hidden genetic variability within a phenotype of multifactorial disease, the phenotype itself can show large differences. The spectrum and the severity of symptoms can vary between the patients even within a single family. This complicates the setting of the affection status in genetic analyses, since prior knowledge of whether this variation is due to genetic or environmental differences is not usually available. Setting of the disease phenotype for the analysis is also complicated if the age at disease onset lies within a large scale; the unaffected family members at the time of the study can develop the disease in later life. For some traits, the availability of quantitative measurements associated with the disease, such as serum cholesterol level in coronary disease or serum IgE levels in asthma, can overcome this problem at least to some extent. Here the phenotype tested is in a quantitative scale, instead of the qualitative, dichotomous affection status where the individuals are classified either affected or unaffected. One should however, bear in mind that the tested single quantitative measure might not correlate fully with the disease itself. Furthermore, it can be difficult to obtain an accurate and comparable measurement of all tested individuals since its value might be strongly dependent on non-genetic factors such as the treatment of the disease and age.

## **DEFINING THE GENOTYPES**

The genetic studies of a given trait need polymorphic markers for identification of the genotype of an individual on the chromosomal region of interest. This genotypic information can be used in association studies between patient and control groups and in linkage approaches in which inheritance patterns of chromosomal regions are followed within families. With family data, the haplotypes, i.e. alleles physically located in the same parental chromosomes can also be constructed if the genotypes of multiple markers in the region are determined.

The first polymorphic markers used were the blood group antigens, followed by serum proteins and HLA antigens. In the 1980's, genetic linkage studies underwent a revolution as the use of restriction fragment length polymorphisms (RFLP) was discovered (Kan and Dozy, 1978). By these dimorphic, single nucleotide polymorphisms (SNPs), which occur randomly and very frequently in the genome with an average distance of 1 kilobase (kb), the genetic studies were not anymore restricted to a few regions. Subsequently, the first human genetic map was constructed for RFLP markers (Donis-Keller et al., 1987). The next type of markers discovered were DNA minisatellites, a variable number of tandem repeats (VNTR) of 14-100 base pair (bp) segments (Bell et al., 1982; Jeffreys et al., 1985). Compared to RFLPs these were more informative for linkage analysis due to multiple alleles, although their large size (>1000 bp) complicates their genotyping.

The most widely used genetic markers today are the DNA microsatellites, another type of VNTR's, also known as single tandem repeats (STR) and simple sequence length polymorphisms (SSLP) (Weber and May, 1989; Litt and Luty, 1989). Microsatellites are very polymorphic tandem repeats of short, 1-6 nucleotide long segments, typically with a high number of alleles differing in the repeat number. They are very frequent and regularly dispersed in the genome; the most widely studied (CA)<sub>n</sub> dinucleotide repeats are found at distances of about 30 kb (Hearne et al., 1992). With the development of polymerase chain reaction (PCR) and automated genotyping technologies, a quick and easy high-throughput genotyping of them has become possible. Due to their high informativity and the availability of dense marker maps for them (Dib et al., 1996), the use of microsatellites has become a standard approach in most genetic analyses today. The mutation rate of microsatellites is higher than other sequences of the genome on average, resulting from slippage during DNA replication which creates alleles with insertions or deletions in the repeat number (Jin et al., 1996). The functionality of microsatellites is unclear, since they are not usually located in the coding regions of the genes. However, some genes contain these repeats in non-coding sequence like 5' or 3' untranslated regions, where in theory they could have an effect on the regulation of mRNA expression or stability. Indeed, expansion of intragenic trinucleotide repeats have been recognized as a major cause certain neurological disorders (Lieberman and Fischbeck, 2000).

The use of single nucleotide polymorphisms have in the past few years become popular again (Nowotny et al., 2001). Although not as informative as microsatellites, SNPs have several advantages. They are far more frequent in the genome and due to a lower mutation rate they are more stable, which is advantageous in linkage disequilibrium based approaches. The wide existence of SNPs in the coding regions of genes also makes them the functional candidate polymorphisms for disease susceptibility. Novel efficient genotyping techniques have recently been developed for SNPs, including the DNA hybridization chips, which would allow extremely large throughput, and the use of SNP panels instead of microsatellites in genomic analyses in the future has been suggested (Kruglyak, 1997).

### **GENOME-WIDE SCREENINGS AND CANDIDATE GENE APPROACHES**

Identification of disease genes for monogenic traits has been successful with a systematic screening of the whole genome by polymorphic markers, followed by the standard linkage analysis in families. For complex diseases the results between independent genome-wide studies have often been contradictory, pointing to the confounding factors complicating the studies of these traits (Risch, 2000). A widely used approach is to first genotype the genome for 350-450 microsatellites. The most interesting regions are often studied further with a denser set of markers and replicated in an independent set of families. Since the average inter-marker distance is – due to practical and economical reasons - as long as 10 centiMorgans (cM), the rate of a false negative result in the genome-wide linkage analysis is high. An alternative is to focus the search on small genomic regions, where plausible candidate genes for the disease are located. The candidates can be e.g. known genes with an assumed functional relevance in the disease pathogenesis, or positional candidate regions suggested by previous genome-wide screenings. In the candidate gene approach, denser sets of markers and larger sets of families can be studied which decrease the chance of a false negative result. However, the genome-wide analysis is the only way to reveal novel regions carrying susceptibility genes, for which no prior assumptions as functional candidates could be made. The significance of the original finding is then easy to test in independent study samples by candidate approaches in these regions.

### **GENETIC LINKAGE ANALYSES**

The term genetic linkage refers to the non-independent segregation of two marker loci in families, i.e. the probability of meiotic recombination is less than 50% (Morgan, 1911). The measure of linkage is the genetic distance or recombination fraction  $\theta$  (theta). When two loci are located so close to each other that no recombinations are observed, the linkage is complete and  $\theta = 0$ . Linkage is absent and  $\theta = 0.5$  when the loci are segregating independently, located far enough from each other or in separate chromosomes. The measure of map distance between

two genetic loci is Morgan (M); 1 cM (centiMorgan) corresponds to  $\theta = 0.01$  in small distances. In longer distances the probability of multiple recombinations between the loci in a single meiosis is higher, and the conversion of recombination fractions to the map distances has to be corrected by Haldane's or Kosambi's map functions. A map distance of 1 cM roughly corresponds to a physical distance of 1 Mb.

### **Parametric linkage analysis**

If the disease parameters such as the inheritance model, disease gene frequency, penetrance and phenocopy frequency are known for the disease, as they usually are for monogenic diseases following the Mendelian dominant or recessive inheritance, the most powerful linkage analysis is the traditional logarithm of odds (LOD) score method (Morton, 1955). The inheritance of marker alleles is followed in large pedigrees with multiple affected individuals. The test calculates the likelihood  $L$  of linkage between the genetic marker and the phenotype with a given recombination fraction  $\theta$ . The observed likelihood is compared to the likelihood of the null hypothesis of no linkage ( $\theta = 0.5$ ). The LOD score is the ten based logarithm of this likelihood ratio. The test uses both genotype and phenotype information of all available family members and the distance between the studied marker locus and the "genuine" risk locus can be estimated. The power of the test is highly sensitive to the disease parameters, which are however usually impossible to accurately estimate for a multifactorial disease.

### **Non-parametric linkage analysis**

For the linkage analysis of complex diseases, methods not relying on the estimates of the disease parameters have been developed. They measure the degree of allele sharing between affected sib-pairs (ASP), or all affected relative pairs (ARP) (Shih and Whittemore, 2001). The observed sharing in families is compared to the sharing probability assuming no linkage. If the marker locus is not linked to the disease locus, the proportions of affected sib-pairs sharing two, one or none of the parental alleles identical by descent (IBD) will not differ significantly from the expected 0.25, 0.50 and 0.25 by chance, if a sufficient number of families are studied. The statistical significance of the sharing excess can be estimated by the  $\chi^2$  test with 2 degrees of freedom. Sharing information can be increased by using markers with a high level of heterozygosity, genotyping several adjacent markers for multipoint linkage analysis and genotyping unaffected siblings especially if one or both parents are unavailable. Current test programs can use estimates of the IBD sharing based on marker allele frequencies and some programs use only identical by state (IBS) sharing information. Among the most widely used non-parametric analyses are the maximum likelihood score method (MLS) and the non-parametric linkage (NPL) which are included e.g. in the MAPMAKER/SIBS and GENEHUNTER program packages (Kruglyak and Lander, 1995; Kruglyak et al., 1996; Markianos et al., 2001).

### **Meta-analysis of genome-wide screenings**

Since the individual genome screenings are usually performed with a limited number of families, because of both availability of families and the high cost of genome-wide mapping, the combined analysis of data from independent genome screenings would be an attractive way to increase the sample size and power to detect risk loci with a modest effect on the disease. This is however, complicated by different marker sets used in the studies, and too often the reluctance of researchers to give access to the raw genotype data. Combining and comparison of the results of individual studies is also problematic, since their magnitude is highly dependent on the study design, i.e. size and pedigree structure of the study sample and the statistical method used. To overcome these problems, several meta-analysis methods have been developed. In genome search meta-analysis (GSMA) the chromosomes are divided into bins of certain length which are then ranked according to their linkage scores (Wise et al., 1999). The bin ranks of each screening study can be then compared and combined. Some type of meta-analysis might prove to be critical for identification of the genes with only a modest effect. On the other hand, combining data showing locus heterogeneity between populations can in fact decrease the overall evidence in regions with a true positive result in some populations. The problem of publication bias towards positive findings is common among studies on complex diseases, which is also important to take into account when collecting data for meta-analysis and interpreting the results.

### **GENETIC ASSOCIATION ANALYSES**

In addition to linkage analyses, the candidate genes can be studied for allelic association. Indeed, the primary susceptibility gene in a genomic region linked with the disease can not be specified by the linkage evidence alone. The linked region, often 10-20 cM in length, can even carry hundreds of genes, among which the one(s) primarily and functionally associated with the trait must be searched for. Instead, the chromosomal region in linkage disequilibrium with the susceptibility gene, and therefore detectable by association analyses, is often only a few centiMorgans. However, as opposed to linkage analysis, the power of association tests is affected by allelic heterogeneity, which may require substantially larger samples to be tested. Two basic approaches are in use: case-control and family-based association studies (Risch, 2000; Cardon and Bell, 2001).

#### **Case-control approach**

Population based association tests compare the allele or genotype frequencies between groups of unrelated patients and independent unaffected controls. The  $\chi^2$  based test is powerful, but the risk of false positive results is high because differences in allele frequencies can also be due to different ethnical background. Careful matching of the control group by age, sex and origin are therefore needed.

### **Family-based association tests**

To overcome the matching problem of case-control studies, methods using family-based controls have been developed (Schaid and Sommer, 1994; Zhao, 2000). The most widely used test is the transmission/disequilibrium test (TDT), a  $\chi^2$  based test for association and linkage, comparing the transmissions vs. non-transmissions of the marker allele from heterozygous parents to affected offspring (Spielman et al., 1993; Spielman and Ewens, 1996). TDT can also be applied for multiple affected siblings but then it is not however, a valid test for association, but can be used as a linkage test. Several extensions of TDT have been developed, which use haplotypic information from multiple markers and sibling genotypes to compensate missing parents. In addition, unaffected controls with parents can be applied to control for spurious significant results due to segregation distortion (Scott and Rogus, 2000).

### **STATISTICAL SIGNIFICANCE**

A traditional criteria for a significant linkage result for single tested locus was a LOD score of  $\geq 3$  ( $p=0.0001$ ) (Chotai, 1984). Since hundreds of markers tested in any single genome-wide analysis increase the chance of false positive result, a genome-wide significance level of LOD 3.3 was suggested (Lander and Kruglyak, 1995). The interpretation of linkage studies on multifactorial traits is however more complicated. The obtained linkage scores do not typically reach these significance levels at all. Different kinds of analyses are often performed on the same data in order to ensure not overlooking linkage, which further decreases the significance due to multiple testing corrections. Therefore, the interpretation of the linkage results of a complex disease is not whether the evidence for linkage is definitive or not, but rather whether it is promising enough for follow-up studies, e.g. LOD above 1 or 2. The succeeded replication of the linkage evidence, even if again only suggestive, in independent samples, can then confirm the true linkage in the region. However, this can not be applied conversely, i.e. failure to replicate the linkage does not necessarily indicate that the original result was false positive, since true heterogeneity for susceptibility genes between the study samples may exist (Risch, 2000).

The significance of association tests often performed for multiple markers carrying multiple alleles may also need to be corrected. However, the Bonferroni correction, where the nominal significance is multiplied with the number of tested alleles or markers is very conservative. This is especially the case for neighboring tightly linked markers for which the allelic associations may not be independent. The computational progresses among genetic tests have offered many sophisticated simulation based approaches to cope with the problem of multiple testing (McIntyre et al., 2000).

## **GENOMEWIDE SCREENINGS AND CANDIDATE GENE APPROACHES IN CELIAC DISEASE**

### ***GENOME-WIDE ANALYSES***

In order to identify novel non-HLA susceptibility loci to celiac disease, five independent genome-wide screenings and several replication studies have been reported so far, including the studies in this thesis. The results are compared further in Table 6 of Discussion. Despite the strong linkage to the HLA region, the findings in other regions are only suggestive and partly controversial between the studies.

The first genome scan in the western counties of Ireland analyzed 15 families with 40 affected sibs (Zhong et al., 1996). In addition to the HLA region which demonstrated strong linkage to CD, six other regions were reported to yield suggestive evidence for linkage including 6p (non-HLA), 7q31, 11p11, 15q26, 19q and 22cen. Subsequently, the non-HLA regions implicated from the Irish study were re-examined by Houlston et al (1997) and Brett et al (1998) using independent sample sets. While the former study showed moderate linkage evidence in the 15q26 region, the latter study did not find any linkage evidence for these regions. The region 15q26 is of particular interest since it harbors a susceptibility gene (IDDM3) for type I diabetes as well (Field et al., 1994). We tested this region for linkage to celiac disease in publication II of this thesis.

The second study analyzed 39 Italian sib-pairs genome-wide and a larger sample of 110 sib-pairs for specific regions (Greco et al., 1998). Suggestive evidence for linkage was found on chromosome 5q. The analysis of a subset of 39 sib pairs in which both siblings manifested a symptomatic form of CD, provided suggestive evidence for linkage to chromosome 11q as well. Interestingly, this finding was not observed in the remaining 71 sib pairs where one sib displayed symptomatic CD and the other had the silent form of CD. Recently, by analyzing an independent sample set, Greco et al (2001) have presented further evidence in support of linkage to chromosome 5q but not to chromosome 11. In study I we investigated the linkage in 5q and 11q candidate regions in Finnish families with CD.

The third genome screening with 16 multiplex UK families (King et al., 2000) reported putative linked loci at 10q and 16q and supporting evidence for linkage to 6q, 11p and 19q, three suggested regions in the first genome scan from Ireland. Lower LOD scores were also found in several other regions. The follow-up study on 17 candidate regions supported linkage at 6p12, 11p11, 17q12, 18q23 and 22q13 (King et al., 2001).

The recent fourth study with Swedish and Norwegian families revealed nominally significant linkage in 8 non-HLA regions, 2q11-13, 3p24, 5q31-33, 9p21, 11p15, 11q23-25, 17q22 and

Xp11 (Naluai et al., 2001). The whole genome analysis was performed on 70 families with affected sib-pairs and the linked regions together with 5 other candidate gene regions were studied with a larger set of 106 families. The stratification of families according to the HLA risk allele dose strengthened the linkage evidence in the CTLA4 candidate gene region at 2q33.

Finally, we studied 60 Finnish families genome-wide in publication V with subsequent follow-up of three selected regions by additional markers and families.

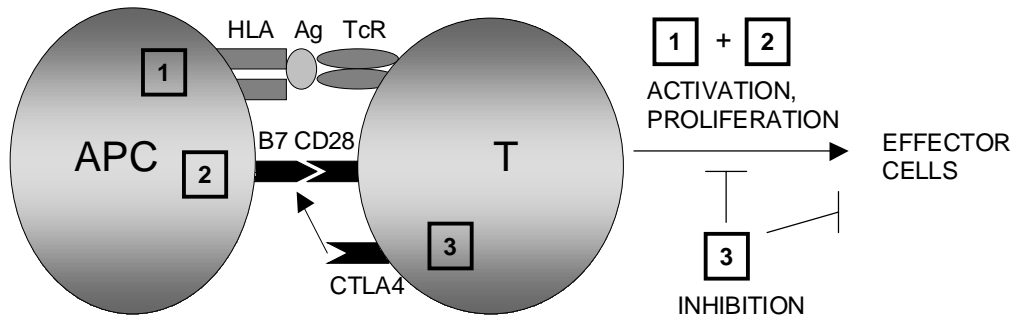
### **CANDIDATE GENES**

Several functional candidate genes have been studied for an association or linkage with celiac disease. The studies are summarized in Table 5 of Discussion. Most of the candidates are players in the immune system that can be assumed to be involved in CD. The other candidates are the genes related to digestive processes and antigen modification, the gene encoding the autoantigen tTG and genomic regions involved in CD-associated diseases. Up to now, the only functional candidate region with increasing evidence of linkage or association with CD is the chromosome 2q33 carrying genes for cytotoxic T-lymphocyte associated antigen-4 (CTLA4), CD28 and inducible costimulator (ICOS). These are all important regulators of T-lymphocyte activation and therefore relevant candidates of CD with T lymphocyte mediated pathogenesis.

### **CANDIDATE GENES CD28, CTLA4 AND ICOS ON CHROMOSOME 2q33**

The antigen-specific signal needed for T-lymphocyte activation is delivered through the binding between the MHC molecule on the antigen presenting cell and the T-lymphocyte receptor on T-lymphocytes. The subsequent immune activation for the antigen requires however, a second signal through the costimulatory molecules such as B7 and CD28 (Lenschow et al., 1996; Bugeon and Dallman, 2000). B7-1 (CD80) and B7-2 (CD86) are transmembrane molecules expressed on professional antigen presenting cells (Figure 5). Their ligand CD28 is constitutively expressed on naive, resting T-lymphocytes. Together with the antigen specific signal, the ligation of B7 molecules with CD28 induces the T-lymphocyte proliferation and effector functions by enhanced expression of T-lymphocyte activating cytokine interleukin-2 (IL2), its receptor and anti-apoptotic molecules. Activation of T-lymphocytes also induces the cell membrane expression of CTLA4 molecule, which shares high homology with CD28 and binds B7-1 and B7-2 with much higher affinity. In resting T-lymphocytes, CTLA4 can only be found intracellularly, but after transportation into the cell membrane during the activation, CTLA4 starts to compete with CD28 for the B7 ligands. Therefore, CTLA4 confers an inhibitory effect on the lymphocyte activation and proliferation. The negative regulatory effect is thought to be due both to the competitive inhibition and to the negative signaling cascade induced (Thompson and Allison, 1997; Lee et al., 1998). The crucial role of CTLA4 in controlling autoreactivity has been

demonstrated by CTLA4 knock-out mice suffering from a massive lethal lymphoproliferation (Tivol et al., 1995; Waterhouse et al., 1995). The role of a recently characterized new member of the CD28/CTLA4 family, inducible costimulator (ICOS), in the immune system is largely unknown, but it seems to be a regulator of Th2 type response and antibody formation (Linsley, 2001).



**Figure 5.** T-lymphocyte activation and co-stimulation. In addition to the antigen specific signal through the HLA-antigen complex and the T lymphocyte receptor (1), the activation of T lymphocyte requires co-stimulatory signals. One of them is delivered by B7-1 or B7-2 receptors expressed by professional antigen presenting cells (APC) and their CD28 ligand on T lymphocyte (2). This two-signal activation allows the proliferation of the T lymphocyte clone and maturation of the effector Th1 or Th2 cells, which can subsequently drive the cell-mediated or humoral immune responses, respectively. The activation process is down-regulated by the increased cell surface expression of the CTLA4 molecule which has a higher affinity to the shared B7 ligands than CD28 (3).

The genes for CD28, CTLA4 and ICOS are located adjacent to each other on chromosome 2q33 (Ling et al., 2001). This region has been found to be linked to several autoimmune diseases including type I diabetes, CD, multiple sclerosis (MS) and autoimmune thyroid diseases (Kristiansen et al., 2000). The CTLA4 gene, the most widely studied candidate of this region, has three known genetic polymorphisms (see Figure 7 in results), which in several studies have been associated with autoimmune diseases. Two of them are single nucleotide polymorphisms: -318\*C/T in the promoter region (Deichmann et al., 1996) and +49\*A/G in the first exon coding for leader peptide of the protein, which leads to Thr -> Ala substitution of codon 17 (Nisticó et al., 1996). A polymorphic AT-repeat region – microsatellite CTLA4(AT)<sub>n</sub> – is located in the 3' untranslated region of the last exon (Polymeropoulos et al., 1991). Any of these polymorphisms could be assumed to alter the expression level or function of CTLA4. Indeed, *in vitro* evidence of decreased inhibitory function of CTLA4 has recently been shown for allele +49\*G (Kouki et al., 2000) and for CTLA4(AT)<sub>n</sub> alleles with long repeat regions (Huang et al., 2000). However, the ultimate identification of primary functional polymorphism within a small segment of DNA can readily be complicated by linkage disequilibrium. In publication IV, we studied the level of LD between the three polymorphisms of the CTLA4 gene.

Evidence for a novel susceptibility locus at 2q33 has also been presented for celiac disease. A French case-control study by Djilali-Saiah et al (1998) presented the first evidence for the CTLA4 association. The frequency of allele A of +49\*A/G polymorphism was significantly higher among celiac patients than in the control group. In study III of this thesis, we found evidence for genetic linkage between seven markers at 2q33 and celiac disease in 100 Finnish CD families.

## **AIMS OF THE STUDY**

In the present thesis the genetic susceptibility to celiac disease was studied in Finnish multiplex families, with the following specific aims:

1. To evaluate the genetic linkage and association of the candidate regions 2q33 (CTLA4/CD28), 5q, 11q and 15q26 with CD (I-III) and to determine the strength of linkage disequilibrium within the CTLA4 gene (IV).
2. To study the presence of linkage heterogeneity due to disease phenotypes and sex in the candidate regions (I).
3. To evaluate the role of HLA-DQ2 in the studied families, with a specific respect in the effect of the dose and sex on the susceptibility to CD (I, V).
4. To screen the whole genome for novel susceptibility loci for celiac disease (V).

## MATERIALS AND METHODS

### STUDY ETHICS

The families studied in this thesis were collected on a voluntary basis through the Finnish Coeliac Society by advertising in the patients' newsletter. The study protocol has been accepted by the Ethical Review Board of Tampere University Hospital (permission number 95173).

### STUDY SUBJECTS

The recruited 137 volunteering families with at least two affected members were accepted for further evaluation. The earlier diagnoses were re-evaluated by scrutinizing the medical records. Healthy family members were screened for antiendomysium antibody positivity and 6.2% of them were found to have an asymptomatic form of the disease (Mustalahti et al., 2002). Families with at least two affected siblings were used in the linkage studies of this thesis (I-III, V). Table 1 shows the number of families and patients in these studies, according to the number of available affected siblings and parents studied. These samples were overlapping sets from a total of 103 families; one of the 100 families in study III was excluded from the other studies due to uncertain identity of one DNA-sample. Haplotypes in study IV were determined from all available founder individuals in a total of 156 families; in addition to the 102 families overlapping with studies I-III and V, 54 additional families were studied. These also included families collected for earlier studies and are described in detail by Polvi et al (1996). Among the 98 families in study V, the diagnosis based on a small bowel biopsy was obtained for all but 17 of the 256 patients, for which no definitive evidence from the biopsy was available. 35 patients had dermatitis herpetiformis and 20 were asymptomatic patients found by screening the family members. Median age at the time of diagnosis was 37 years. All families were of apparent Finnish origin and there was no evidence of any particular clustering in their current places of residence.

**Table 1.** Structure of Finnish families with at least one affected sib-pair studied for publications I-III and V. The family samples represent overlapping sets from a total of 103 families.

Study	Number of families	Affected siblings			Available parents			Total of patients
		2	3	4	2	1	0	
I	102	71	24	7	42	27	33	263
II	99	69	24	6	39	27	33	253
III	100	75	20	5	39	28	33	250
V (stage 1)	60	39	16	5	19	20	21	159
V (stage 2)	98	67	24	7	42	24	32	256

## GENETIC MARKERS

All polymorphic microsatellites were genotyped with fluorescence based detection systems of PCR-products using ABI 310 and 377 Sequencers (Applied Biosystems, Foster City, CA). Microsatellites genotyped in candidate region 2q33 (III) were D2S2392, D2S2214, D2S116, CTLA4(AT)<sub>n</sub>, D2S2189 and D2S2237, at 5q (I) D5S410, D5S422, D5S2032, D5S425, D5S2069 and D5S2111, and at 11q (I) D11S898, D11S4111, D11S4142, D11S976, D11S4171, CD3D (Mfd69CA, dinucleotide repeat within CD3D gene), D11S934 and D11S910. Two SNPs within the CTLA4 gene (III-IV) were genotyped by the PCR/RFLP method using the primers 5'TTACGAGAAAGGAAGCCGTG, 5'AATTGAATTGGACTGGATGGT and *Mse*I-digestion for CTLA4-318\*C/T and the primers 5'AACCCAGGTAGGAGAAACAC, 5'GCTCTACTTCCTGAAGACCT and *Bbv*I-digestion for CTLA4+49\*A/G polymorphism, with subsequent agarose gel electrophoresis. Genotypic data in the candidate regions were checked for Mendelian errors and for unexpected double recombination events within small genetic distance, using multipoint analysis of Genehunter-plus v1.1 or Genehunter 2.0 program versions (Kruglyak et al., 1996). The genotypes most likely to be erroneous were either retyped or removed from the analyses. The genome-wide screening (V) was performed with 352 microsatellites with an average marker distance of 9.6 cM, and the genotypic data was examined for possible error sources in Mendelian consistency and biological relationships using the Pedcheck (O'Connell and Weeks, 1998) and the RelCheck computer programs (Broman and Weber, 1998), respectively. Primer sequences for all microsatellites (except the CTLA4(AT)<sub>n</sub>) and the genetic distances between the markers were obtained from public marker databases of Genethon ([www.genethon.fr](http://www.genethon.fr)), the Genome Database ([www.gdb.org](http://www.gdb.org)), and the Marshfield Genetic Laboratory ([research.marshfieldclinic.org/genetics](http://research.marshfieldclinic.org/genetics)). The genotyping of the marker CTLA4(AT)<sub>n</sub> was performed using the primers 5'GTGATGCTAAAGGTTGTATTGC and 5'AAAACATACGTGGCTCTATGCAC.

## HLA TYPING

The HLA-DQB1 alleles were genotyped from all available family members using a DQ SSP "Low Resolution" kit (Dynal AS, Oslo, Norway). Major classes of the HLA-DQB1 alleles, including the known susceptibility alleles DQB1\*02 (DQ2) and DQB1\*0302 (DQ8) could be determined. Based on our unpublished results, about 90% of DQB1\*02 positive individuals in Finnish families with celiac disease have the DQA1\*05 DQB1\*02 haplotype.

## DATA ANALYSIS

The linkage in the candidate gene regions (I-III) was tested with the Genehunter-plus v1.1 and the Genehunter 2.0 program versions (Kruglyak et al., 1996). Multipoint NPL<sub>all</sub> or MLS scores

were reported. The heterogeneity between the MLS results of the phenotypical subgroups of the families (I) was tested using the M-test (Morton, 1956), which compares the linkage results in the subgroups a and b with the linkage in the whole sample (a+b) as  $\chi^2$  statistics  $2\ln(10)$  [MLS(a)+MLS(b)-MLS(a+b)]. The significance of the test result was assessed by 10000 replicates of randomized division of the whole sample in subgroups with sizes equal to the original subgroups, with following linkage and M-tests. The proportion of these simulated test results exceeding the observed value was set as the simulated p-value.

The transmission/disequilibrium test (TDT) for the candidate regions was performed by the Genehunter 2.0 (I-II) and Sib-Pair 0.95.3 (III) programs (Kruglyak et al., 1996; Duffy, 1997). For a valid association test, only one affected offspring in each family was tested (I-II). For both one and two locus TDT, the significance of the test result was obtained by the permutation option in Genehunter 2, to correct the effect of the testing of multiple markers with multiple alleles (I-II).

Allelic association in the publication III was tested in 100 families using the Sib-Pair 0.95.3 version (Duffy, 1997). Allelic frequencies were compared between all affected and unaffected family members by the Pearson goodness-of-fit based test. To correct the bias in nominal P-values due to relatedness of individuals within families, an empirical P-value was estimated by the Monte-Carlo simulation using 10000 iterations.

Linkage disequilibrium was calculated by the Arlequin program (Schneider et al., 1997) for a given set of independent three-locus haplotypes in the publication V. Haplotypes were constructed by the Genehunter 2.0. An exact test of LD between 3 pairs of markers and the standardized disequilibrium value  $D'$  for each of the allele pairs among the marker pairs were calculated. Values of  $D'$  range from  $-1.0$ , stating that the alleles never occur together, to complete disequilibrium  $1.0$ , through  $0$  stating for no LD between the alleles.

Genome-wide linkage analysis (V) was performed using a pseudomarker approach (Göring and Terwilliger, 2000a; Göring and Terwilliger, 2000b) by the FASTLINK program (Cottingham Jr et al., 1993). Two dominant and two recessive models of inheritance were tested, one with high penetrance (0.90) and the other with very low penetrance (0.0007). No phenocopies were allowed and the disease allele was assumed to be infinitesimally rare, according to the pseudomarker strategy (Göring and Terwilliger, 2000b). Multipoint analyses of the three follow-up regions of the genome scan were conducted by joining the meiotic information from two neighboring markers (i.e. the three-point analysis), using only the genetic model yielding the highest lod score in the two-point analysis of the region. Multipoint  $NPL_{all}$  statistics for the 60 families were calculated genome-wide by Genehunter 2.0 (unpublished data).

## RESULTS

### LINKAGE IN CANDIDATE REGIONS 2q33, 5q, 11q AND 15q26 (I-III)

The candidate region 2q33 harboring the CTLA4, CD28 and ICOS genes was studied since it shows association or linkage to several autoimmune diseases (Kristiansen et al., 2000) and at the time of the study, an allelic association with celiac disease was also reported (Djilali-Saiah et al., 1998). The studied chromosomal regions 5q and 11q were the candidates suggested in the genome-wide analysis by Greco et al (1998). The region 15q26 was studied as it was reported to be linked with CD in two previous studies (Zhong et al., 1996; Houlston et al., 1997), and may also confer genetic susceptibility to type I diabetes (Field et al., 1994).

The highest non-parametric linkage results, either MLS or  $NPL_{all}$  scores with nominal p-values, for the four studied candidate regions are summarized in Table 2. Details of the studied markers are shown in the original publications I-III. In the total set of families, suggestive evidence for linkage was found for 5qter ( $p=0.03$ ), 11q23 (0.01) and 2q33 ( $p=0.006$ ). On the other hand, no evidence for linkage between 15q26 and celiac disease was observed.

**Table 2.** Linkage between celiac disease and the candidate gene regions 2q33, 5q, 11q, 15q26 in Finnish families with affected sib-pairs.

Region	Families studied	Markers studied	Length of the region (cM)	Marker with the highest score	Test statistics	p-value (nominal)
2q33 (III)	100	7	3.3	D2S116	$NPL_{all}$ 2.55	0.006
5q (I)	102	6	34.5	D5S2111	MLS 0.88	0.03
11q (I)	102	8	48.3	D11S4142	MLS 1.37	0.01
15q26 (II)	99	5	19.9	D15S642	$NPL_{all}$ -0.07	0.53

### ASSOCIATION AND TDT RESULTS ON CANDIDATE GENES

At 2q33, an association was found between CD and marker D2S116 located 0.3 cM centromeric to CTLA4. Allele 136 bp was present in 5% (23/476) of the chromosomes of all patients compared to 12% (19/162) among all genotyped healthy individuals ( $\chi^2$  9.35,  $p=0.0022$ ) and the difference remained significant after Bonferroni correction for 8 tested alleles of D2S116 ( $p=0.018$ ). Association of this marker with celiac disease was significant ( $p=0.0001$ ) also with the simulation approach which corrects the relatedness of the tested family members. The same allele D2S116\*136 showed negative maternal transmission with the transmission/

disequilibrium test (1 transmitted versus 11 non-transmissions, nominal  $p < 0.05$ ). None of the other six tested markers at 2q33 showed association with CD.

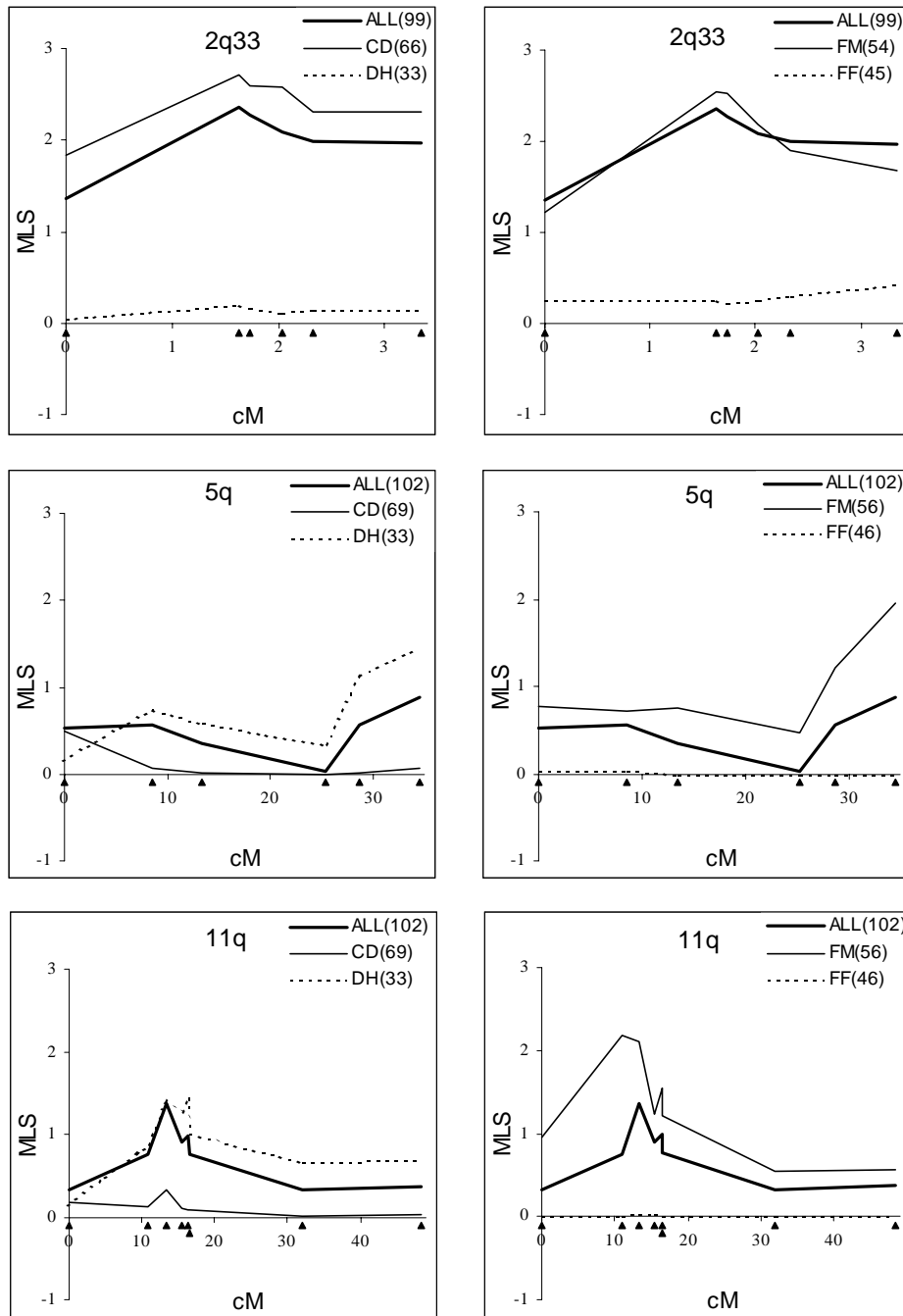
No evidence for allelic association was found for markers on 5q. Instead, TDT supported linkage between CD and the 5 adjacent markers in region 11q23. In the single locus test, allele D11S4171\*131 showed preferential transmission to affected offspring (21 transmitted versus 7 not transmitted, nominal  $p = 0.008$ ) and nominally significant scores were obtained for all four adjacent marker pairs. The haplotype D11S4171\*131-CD3D\*82 also showed evidence of allelic association with CD when only one affected sib from each family was analyzed. However, none of the TDT statistics remained significant after the permutation approach to correct for multiple testing.

While the NPL analysis for markers at 15q26 revealed no evidence for linkage with CD, a preferential transmission of the allele D15S107\*1 for the affected offspring was found (20 transmissions vs. 5 non-transmissions, nominal  $p = 0.003$ ). The difference remained significant after the permutation test as well (0.03). In the two-locus TDT, the haplotype D15S107\*1-D15S120\*6 was transmitted 14 times, compared to 2 non-transmissions (nominal  $p = 0.003$ , permuted  $p = 0.058$ ). No particular shared feature in HLA-DQ status, geographical origin, or disease phenotype was found for families carrying this haplotype.

## LINKAGE HETEROGENEITY (I)

Heterogeneity for linkage results between the phenotypic subgroups of the families was tested in the candidate regions 2q33, 5q and 11q. Two divisions could be made, still maintaining sufficiently large subgroups for linkage analysis: disease manifestation and sex of the affected siblings. For this, linkage was tested separately in 33 families with at least one patient affected by dermatitis herpetiformis (DH) and in the remaining 69 families with no DH patients, and in 56 families with at least one affected male sibling and in 46 with only affected female sibs. Figure 6 shows the results which varied widely between the subgroups. Linkage to 2q33 was only found in families with no DH patients (CD group) and conversely, the families with DH patients showed linkage to the 5q and 11q markers (DH group). In all three regions, linkage was detected in those families with at least one male patient (FM group), whereas the family group with only females affected (FF group) showed no linkage to these regions. The only statistically significant difference with the M-test between the subgroups was for the FM and FF division at 5q (simulated  $p = 0.019$ ). The linkage at HLA region was also tested with the DQB1 genotype data. Extremely strong linkage with MLS of 14.3 ( $p = 10^{-16}$ ) was observed in a total sample of 102 families, and the linkage was highly significant in all the tested subgroups, MLS varying from 3.8 to 10.9.

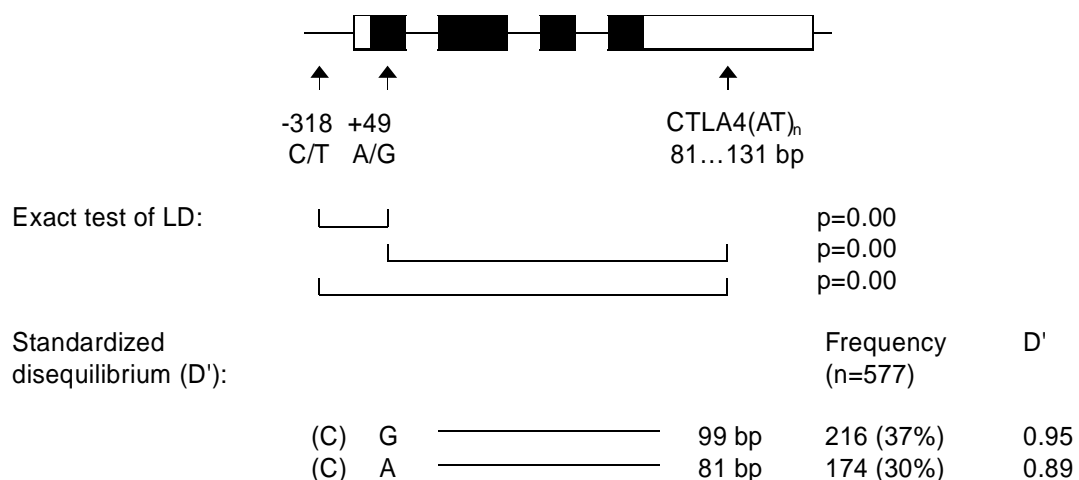
RESULTS



**Figure 6.** MLS scores in the candidate regions 2q33, 5q and 11q in the whole group of families and in the subgroups. A total set of 99 (2q33) or 102 (5q, 11q) families (group ALL) were divided into subgroups DH and CD, based on the presence or absence of patients with dermatitis herpetiformis in the families, respectively. Another division for FM and FF groups was based on the presence or absence of male patients among the affected siblings in the families, respectively. Only 10 families contained male patients with no affected females, and they were included in the FM group.

## LINKAGE DISEQUILIBRIUM WITHIN THE CTLA4 GENE (IV)

Two previous studies showed that variations in the CTLA4 gene may affect the inhibitory function of CTLA4 in the activation of T-lymphocytes (Kouki et al., 2000; Huang et al., 2000). As these authors concentrated on two separate variations, we studied the strength of linkage disequilibrium between all three known polymorphisms within the gene. LD was tested among 577 independent haplotypes constructed from a total of 156 non-related Finnish families with celiac disease (Figure 7). An exact test of linkage disequilibrium showed a highly significant level of LD between all three marker pairs ( $p=0.00$ ). Two common haplotypes were observed: -318\*C; +49\*G; CTLA4(AT)<sub>n</sub>\*99 bp with a frequency of 37% and -318\*C; +49\*A; CTLA4(AT)<sub>n</sub>\*81 bp with a 30% frequency, showing a high level of LD between the alleles at +49\*A/G and CTLA4(AT)<sub>n</sub> ( $D' = 0.95$  and  $0.89$ , respectively). The other allelic combinations showed much lower frequencies (0.2 – 6.0%) and are presented in Table I of the original publication IV.



**Figure 7.** Linkage disequilibrium between the CTLA4 gene markers. Genomic organization of the CTLA4 gene (not in scale) is shown as ■ for coding regions and as □ for the untranslated regions of exons. The relative positions of the three known polymorphisms are marked with arrows. P-values of the exact test of LD between marker pairs, and D' values and frequencies for the two most common allele combinations are shown.

RESULTS

**GENOME-WIDE LINKAGE STUDY (V)**

Figure 8 shows the pseudomarker LOD scores of all the chromosomes tested over the four inheritance models in the 60 families affected by CD. Highly significant linkage for all the models was found on chromosome 6p21 overlapping the HLA region (LOD scores between 4.63 and 5.96 for the marker D6S1281). Six non-HLA regions 1p36, 4p15, 5q31, 7q21, 9p21-23 and 6q12 showed suggestive evidence for linkage with maximum LOD scores over 1.00 (1.02 - 2.11), as presented in Table 3. A follow-up study with a denser marker set and 38 additional CD-families was performed for three of these regions: 4p15 was selected due to its highest LOD score of 2.11, and 5q31 and 7q21 due to the reported linkage evidence close to these regions in the two previous genome-wide studies. These regions were analyzed using the model which yielded in the best linkage score in the genome-wide analysis, and the two- and three-point results for the original set of 60 and the combined set of 98 families are shown in Table 3. Although the three-point analysis with dense marker set yielded stronger evidence for linkage at 4p15 (LOD 3.25) and consistent evidence at 5q31 and 7q21, the addition of families did not strengthened the evidence in any of the three regions.

**Table 3.** LOD scores over 1.0 obtained for the six non-HLA regions in the genome-wide analysis and in the follow-up study of three of these regions. Only the highest scores among the four inheritance models are shown; either with dominant or recessive inheritance for a gene with high or low penetrance. Multipoint  $NPL_{all}$  scores with nominal p-values in the same six regions are presented for comparison, as well as the linkage results for other regions of interest: two suggested in studies I and III and two reaching nominally significant linkage only with  $NPL_{all}$  statistics. From the follow-up analysis, the results from markers yielding the highest linkage scores are reported. nd, not determined.

Locus	Marker	Model	Genome-wide analysis 60 families		Follow-up analysis			
			Two-point LOD	Multipoint $NPL_{all}$ (p)	60 families Two-point	60 families Three-point	98 families Two-point	98 families Three-point
<b>REGIONS WITH LOD&gt;1</b>								
1p36	D1S3669	Dom/High	1.09	1.60 (0.06)	nd	nd	nd	nd
4p15	D4S2639	Dom/Low	2.11	2.33 (0.01)	2.14	3.25	1.78	2.40
5q31	D5S816	Rec/High	1.72	1.47 (0.07)	1.72	1.50	1.81	1.52
7q21	D7S821	Dom/High	1.02	1.09 (0.14)	1.01	1.04	0.84	1.02
9p21-23	D9S741	Rec/Low	1.11	1.76 (0.04)	nd	nd	nd	nd
16q12	D16S3253	Rec/Low	1.40	1.79 (0.04)	nd	nd	nd	nd
<b>OTHER REGIONS OF INTEREST</b>								
11q (I)	D11S2002	Dom/High	0.20	0.93 (0.18)	nd	nd	nd	nd
2q33 (III)	D2S1361	Dom/Low	0.47	1.30 (0.10)	nd	nd	nd	nd
4q26	D4S3250	Dom/Low	0.54	2.14 (0.02)	nd	nd	nd	nd
15q11	D15S165	Dom/Low	0.82	2.18 (0.02)	nd	nd	nd	nd

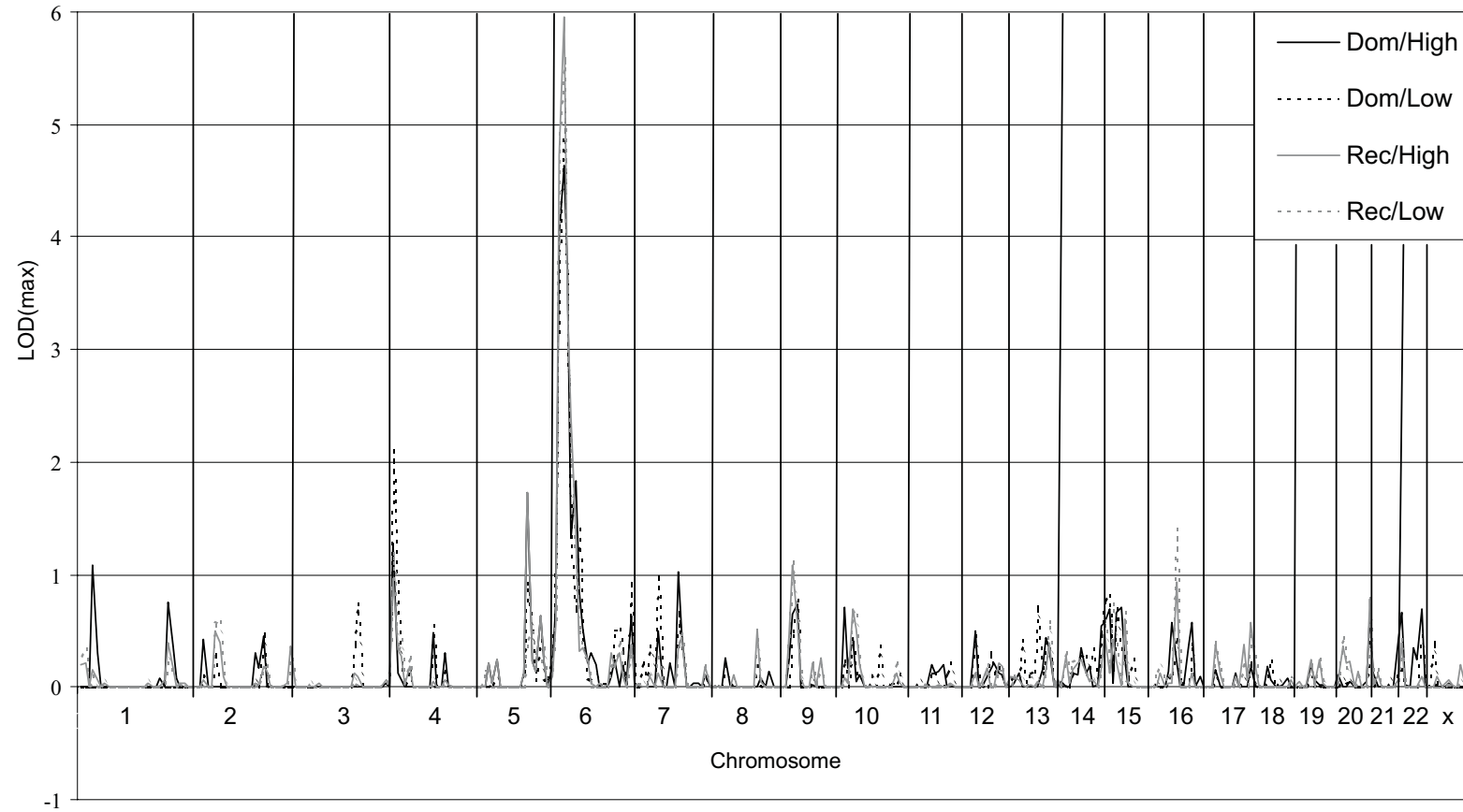


Figure 8. Genome-wide linkage in the 60 celiac disease families. Maximum LOD scores are shown for the four tested models with either high or low penetrance and dominant or recessive inheritance.

## RESULTS

Table 3 also shows the highest multipoint  $NPL_{all}$  statistics for the suggested six non-HLA regions in 60 families (unpublished data). The results were consistent with the LOD scores, varying from  $NPL_{all}$  1.09 to 2.33. Only markers on chromosomes 4p, 9p and 16q showed nominally significant ( $p < 0.05$ ) results. Both statistics are also shown for four other interesting regions. 11q and 2q33 showed linkage in the candidate studies I and III, but no linkage was observed in the genome-wide analysis. Chromosomal regions 4q26 and 15q11 reached nominally significant  $NPL_{all}$  results but remained undetected in the pseudomarker analysis with LOD scores under 1.0.

### SUSCEPTIBILITY TO CELIAC DISEASE DUE TO HLA FACTORS (I, V)

All available members of the 102 studied families were genotyped for the HLA-DQB1 alleles. Our unpublished results have shown that over 90% of the DQB1\*02 positive patients in Finland carry the common DRB1\*03 – DQA1\*05 – DQB1\*02 risk haplotype (often referred to as DR3-DQ2). For simplification, all DQB1\*02 positive were therefore considered as DQ2 risk molecule carriers, in the absence of data for DQA1 alleles at the time of the studies. Among the 98 families used for the genome-wide scan (V), 252 of the 260 patients carried DQ2, seven were negative for DQ2 but carried the DQ8 risk molecule and only one patient was negative for both DQ2 and DQ8.

Table 4 shows the disease prevalence among all DQ2 positive family members, as well as among those with either a single or double dose of DQ2. In addition, division by sex is shown. 65% of all DQ2 carriers were affected by celiac disease. The CD prevalence was higher (79%) in individuals with two copies of DQ2, compared to the heterozygous carriers (60%). Since the individuals tested were not genetically independent, testing of the statistical significance for this difference, suggesting dose dependence of risk due to DQ2, can not be regarded as valid. CD prevalence among DQ2 carriers also differed between the sexes. 71% of all DQ2 positive females in the families were affected, compared to only 52% of males (Fischer exact  $p = 0.0001$ ). The tested individuals were genetically related in this comparison as well, but as the sex of patients is not dependent on parental HLA genotypes, the test result can be regarded as suggestive evidence for sex-dependent risk of CD due to HLA-DQ2.

**Table 4.** Prevalence of CD among all family members due to the HLA-DQ2 risk allele. The dose effect of DQ2 was studied in 98 families (V) and the sex-specific risk among DQ2 positive family members in overlapping 102 families (I). \* $p=0.0001$

	Affected	Unaffected	CD prevalence
<b>Overall prevalence</b>			
All DQ2+	252	137	65%
<b>Dose effect</b>			
DQ2 single dose	175	117	60%
DQ2 double dose	77	20	79%
<b>Sex effect</b>			
DQ2+ females	176	71	71%*
DQ2+ males	84	79	52%*

## DISCUSSION

In the studies of this thesis, we investigated genetic susceptibility to celiac disease due to HLA-unlinked genes. We concentrated on the functionally interesting CTLA4/CD28 gene region on chromosome 2q33 (III) and on the previously suggested candidate gene regions 5q, 11q and 15q26 (I-II). The whole genome was screened for novel susceptibility loci for CD (V). In addition, linkage heterogeneity for the candidate gene regions was further studied between the disease phenotypes and sex, and the sex- and dose-related roles of HLA risk factors in celiac disease were investigated (I, V).

### CANDIDATE GENES IN CELIAC DISEASE

Several linkage and association studies have been performed for HLA-unlinked functional candidate genes of CD (Table 5). The results of the five genome-wide screenings and subsequent replication studies are summarized in Table 6. Among the functional candidates involved in immune responses are the genes encoding the T-lymphocyte activation regulators CTLA4, CD28 and ICOS, their ligands B7-1 and B7-2, T-lymphocyte receptor (TcR), the CD4 antigen and the GM allotypes encoded by the genes for immunoglobulin heavy chain constant regions. Other studied functional candidates are the genes for the CD autoantigen tissue transglutaminase (tTG), two ectopeptidases dipeptidylpeptidase IV (DPP4) and aminopeptidase N (APN) involved in digestive processes, and the candidate gene regions due to CD-associated diseases, such as the susceptibility loci in type I diabetes and the chromosomes 7q11 and 21 involved in Williams' and Down Syndromes, respectively. The 2q33 region carrying genes for CD28, CTLA4 and ICOS can be regarded as the most promising functional candidate region in CD. Instead, evidence in other regions is poor, despite weak evidence for the TcR- $\gamma$  gene (Arai et al., 1995) and GM allotypes (Bouguerra et al., 1999). It should be noted however, that most of the older studies were case-control approaches with relatively small sample sizes. In these studies the chance of false positive or negative result is likely to be higher than in the family-based linkage and/or association studies. The family-based studies have been marked with asterisks in the Table 5.

At the time of study III, an association between celiac disease and allele CTLA4+49\*A had been reported by a case-control approach performed in the French population (Djilali-Saiah et al., 1998). We studied both family-based linkage and association for seven markers at 2q33 in a total of 100 multiplex families (III). Linkage in this region was statistically significant ( $p=0.006$ ). However, no allelic association for the two studied CTLA4 gene polymorphisms could be detected, and therefore the association of CTLA4+49\*A in the French population was not confirmed in our study. Instead, allelic association was observed with marker D2S116 located close to the CTLA4 gene. The results indicate that the gene region may carry a genuine risk

factor of CD. This has also been supported by two other recent studies in Scandinavian families which interestingly also showed association with the CTLA4+49\*A allele (Torinsson-Naluai et al., 2000; Popat et al., 2002). Another study with British families also showed linkage at 2q33 (King et al., 2002). As in our study, no association was found with CTLA4 gene markers, but instead with D2S2214, a microsatellite marker adjacent to the D2S116 and CTLA4 gene. It is of note that the possible association between allele A at CTLA4+49 marker and celiac disease differs from other autoimmune diseases such as type I diabetes, autoimmune thyroid diseases and rheumatoid arthritis in which association with allele G has been frequently observed (Kristiansen et al., 2000). This suggests that the susceptibility factor in this gene region might differ between celiac disease and the other disorders, although the primarily associated gene and polymorphism still remains to be determined for all of them.

Although the evidence for a susceptibility factor to autoimmune diseases on chromosome 2q33 is clear, many discrepancies of linkage evidence or associated alleles still exist between the studies. The lack of linkage may indicate either true heterogeneity between the studied

**Table 5.** Studies on HLA-unlinked functional candidate genes in celiac disease. \*Family based studies for linkage and/or association.

CANDIDATE GENE REGION	POSITIVE EVIDENCE	NO EVIDENCE
2q33 CTLA4, CD28, ICOS	Djilali-Saiah et al (1998) Holopainen et al (1999)* <b>(III)</b> Torinsson-Naluai et al (2000)* King et al (2002)* Popat et al (2002)*	Clot et al (1999)* Neuhausen et al (2001)*
3q13-21 B7-1, B7-2 7p15-14, 7q35, 14q11 TcR genes ( $\alpha$ , $\beta$ , $\gamma$ or $\delta$ )	Arai et al (1995) TcR- $\gamma$	Ali-Varpula et al (2002)* Roschmann et al (1993)* Roschmann et al (1995)* Roschmann et al (1996)* Yiannakou et al (1999)* Neuhausen et al (2001)*
12pter-12 CD4 14q32 GM Ig-allotypes	Kagnoff et al (1983) Carbonara et al (1983) Weiss et al (1983) Rautonen et al (1990) Caruso (1991a,b) Mazzola (1992) Bouguerra et al (1999)*	Fedrick et al (1985) Hetzl et al (1986) Hannigan et al (1991)
20q11 tTG 2q23 dipeptidylpeptidase IV (DPP4) 15q25-26 aminopeptidase N (APN)		van Belzen et al (2001)* Clot et al (2000)*
15q26(IDD3), 11q23(IDD4), 6q25(IDD5), 18q21(IDD6), 6q27(IDD7), 3q22-25(IDD9), 10p11- q11(IDD10)		Neuhausen et al (2001)* Susi et al (2001)* 15q26 <b>(II)</b>
7q11.23 Williams Syndrome locus chromosome 21 Down Syndrome locus		Grillo et al (2000)* Morris et al (2000)*

## DISCUSSION

populations or false negative results which are likely to be common in multifactorial diseases. Two candidate gene studies on celiac disease have also failed to show linkage or association to 2q33, one performed in Italian and Tunisian families (Clot et al., 1999a) and the other in North American families (Neuhausen et al., 2001). Interestingly, any evidence for linkage on chromosome 2q was found neither in our genome-wide screening (V) nor in the first two genome-wide studies (Zhong et al., 1996; Greco et al., 1998). The strength of linkage in this region was also weak in the genome screen by King et al (2000). The study in Swedish and Norwegian families showed linkage at 2q11-13, a region located as far as 80 cM centromeric from 2q33 (Naluai et al., 2001). Their subsequent stratification analysis suggested linkage at 2q33 in a subgroup of families with a single dose of the HLA-DQ2 risk molecule. However, their earlier finding of allelic association of CTLA4+49\*A with CD (Torinsson-Naluai et al., 2000) was not strengthened in this subgroup. The discrepancy between the genome-wide and candidate gene approaches in chromosome 2q33 can indicate the increase of power in candidate gene studies usually performed with denser marker sets and a larger number of families, which was also the case in our studies. Denser marker sets also allow the application of methods based on linkage disequilibrium. Good examples of the advantages of these methods have been shown for other complex diseases as well. The linkage and association between type I diabetes and its candidate genes for insulin (11p15) as well as for CTLA4, have been successful with TDT-type approaches, while the traditional linkage analyses at these loci have shown discrepancies even with large data sets (Spielman et al., 1993; She and Marron, 1998).

Linkage disequilibrium often complicates the identification of which polymorphism confers the primary functional susceptibility. The best known example is the MHC gene cluster showing very strong LD which can lead to almost equivalent disease associations of several genes present in the same conserved haplotype. In the candidate region 2q33, most of the genetic studies in autoimmune diseases have concentrated on three available polymorphisms in the CTLA4 gene and revealed somewhat discrepant results on the disease-associated polymorphisms. Recently, two studies showed that genetic variation in the CTLA4 gene can have an effect on the downregulatory function of CTLA4 on activated T lymphocytes. Huang et al (2000) studied the CTLA4(AT)<sub>n</sub> microsatellite 3' untranslated region and found a decreased inhibitory function of CTLA4 in individuals carrying alleles with a high number of repeats. A similar effect was found by Kouki et al (2000) in individuals who carried the G allele at CTLA4+49\*A/G polymorphism leading to Ala->Thr amino acid change in the leader peptide of CTLA4. In study IV we showed in a large sample of Finnish haplotypes that linkage disequilibrium between the three known CTLA4 polymorphisms is extremely high. For example, the most frequent allele with high number of repeats CTLA4(AT)<sub>n</sub>\*99bp nearly always occurred with the allele CTLA4+49\*G in a common haplotype with a frequency of 37% in Finland. Our study demonstrated that the identification of the primary functional polymorphism can not be done by studying only one polymorphism at a time. Therefore, either more detailed haplotype analyses



## DISCUSSION

or *in vitro* mutagenesis approaches are needed, since all three CTLA4 polymorphisms are functionally relevant candidates. This is supported by a recent functional study on all three CTLA4 polymorphisms which interestingly showed a differential expression of CTLA4 due to the genotype at the promoter variation (Ligers et al., 2001). As this marker was not included in the studies by Huang et al (2000) or Kouki et al (2000), any conclusions of the primarily functional polymorphism can not be drawn yet. The role of other closely linked and functionally related genes, such as CD28 and ICOS, can not be excluded either. Furthermore, no prior assumption can be made for only one polymorphism or one of these genes to confer susceptibility to autoimmune diseases. Studies on all these genes are therefore justified with specific respect in the linkage disequilibrium between them. We have recently screened the human ICOS gene for polymorphisms to be used in genetic studies (Haimila et al., 2002) and our unpublished results show that LD is significant between these three genes in the Finnish celiac families. Controversially, no LD was found between CD28, CTLA4 and ICOS in the Japanese population (Ihara et al., 2001). This can either result from the differences in the sample type and selected markers between the studies, or from the higher level of LD in the Finnish population featuring a founder effect which points to the relevance of LD measurement in all populations under study in the future.

In addition to the candidate region on 2q33, evidence for the susceptibility loci for CD on chromosomes 5q and 11q is also accumulating. Our results of the study I supported the genetic linkage in these regions as originally suggested by Greco et al (1998). Although these authors could not confirm the linkage at 11q in a larger family sample (Greco et al., 2001), weak evidence for linkage at 11q was found also in studies by King et al (2000&2001). Interestingly, the genome screen on Swedish and Norwegian families supported linkage at 5q and 11q as well (Naluai et al., 2001). The region 11q23 linked with CD in our study carries several functional candidate genes involved in the immune system, such as the interleukin-10 receptor gene and genes encoding three CD3-complex chains. The region also harbors genes for matrix metalloproteinases MMP-1 and MMP-3, expression of which is increased in the untreated celiac mucosa and which may be involved in the pathogenetic process of CD (Daum et al., 1999). In our study, the linkage was also supported by a preferential transmission to patients of certain haplotypes within this gene region. The original report by Greco et al (1998) suggested that this locus could distinguish the symptomatic and silent forms of celiac disease. Unfortunately, this could not be tested in our families due to the small number of families including sibs affected with silent CD. Despite the linkage evidence in our candidate gene approach, no linkage on 11q was detected in our genome-wide study. This can be due to a higher false negative rate in genome-wide studies performed in a restricted number of markers, and in our case also with smaller number of families.

Instead, both our candidate gene study (I) and the genome-wide screening (V), supported the positive linkage on chromosome 5q suggested in two previous genome-wide studies by Zhong et al (1996) and Creco et al (1998). The follow-up study by Greco et al (2001) strengthened the

linkage in the Italian population and in addition, the results of the other Scandinavian genome-wide study by Naluai et al (2001) and an unpublished meta-analysis of four European genome-wide linkage results have supported well the linkage in this region. Although the regions with the strongest linkage to 5q did not completely overlap between these studies, the most interesting functional candidate is the cytokine gene cluster located in 5q31 which harbors the genes for interleukin (IL)-3, IL-4, IL-5, IL-9 and IL-13. These cytokines are mediators of Th2 type immune responses and therefore good candidates for many immune-mediated diseases in which the balance of Th1 and Th2 responses is affected. In addition to celiac disease, this gene region has also been linked to asthma (Xu et al., 2000; Yokouchi et al., 2000; Lonjou et al., 2000; Palmer et al., 2001; Kauppi et al., 2001) and Crohn's disease (Rioux et al., 2001), another chronic autoimmune-type inflammation of the small intestine.

No clear evidence for linkage in Finnish families could be detected either by our candidate approach (II) or by the genome-wide study (V) in the 15q26 region suggested by two studies for populations of the British Isles (Zhong et al., 1996; Houlston et al., 1997). The region is also of particular interest, because of its involvement in type I diabetes, an autoimmune disease associated with CD (Field et al., 1994). However, no linkage to this region has been found in any other recent studies on CD, suggesting either the presence of a true susceptibility locus only in British populations, or false positive findings in these original studies performed with relatively small sample sizes.

In addition to 5q, our genome-wide analysis revealed five other HLA-unlinked regions with a weak to moderate linkage to CD. Interestingly, linkage near to all of these regions have been reported in at least one independent study on CD (Table 6). The regions 1p36, 4p15 and 16q12 maps close to regions with some evidence also in studies by King et al (2000&2001). Linkage at 9p21 was suggested in other Scandinavian populations by Naluai et al (2001). Similarly, the region 7q21 maps close both to 7q31 reported by Zhong et al (1996) and to the Williams' Syndrome (WS) locus at 7q11.23. WS is a rare monogenic disease resulting from a chromosomal deletion around the elastin gene at this locus, and WS patients have an increased risk to develop celiac disease (Giannotti et al., 2001). A previous candidate gene study at this gene locus could not however, find any linkage evidence for CD (Grillo et al., 2000). The regions 4p, 5q and 7q were further followed up with additional markers and 38 families. Despite the high LOD score of 3.25 obtained at 4p15 by multipoint analysis in the original 60 families, nearly reaching the genome-wide significance level of 3.3, the analysis in the total set of 98 families did not markedly bolster the linkage evidence in any of these regions. This may point to the underlying heterogeneity between the original and additional families, or alternatively, to false positive findings. As the linkage evidence in any of these regions is not definite, subsequent confirming studies in independent samples are needed.

## DISCUSSION

Conflictingly, the genome-wide analysis failed to show linkage on chromosomes 11q and 2q which were suggestive in our previous candidate studies I and III. This discrepancy may result from the enhanced power due to the larger family sample and the denser marker set used in these former studies. There may also be slight differences between the analysis methods. Although not significant, the obtained  $NPL_{all}$  scores (0.93 for 11q and 1.30 for 2q) were close to the level of the six regions that exceeded LOD 1.0 with pseudomarker approach (unpublished data). Furthermore,  $NPL_{all}$  analysis revealed two interesting regions on chromosomes 4q and 15q with nominally significant linkage ( $p=0.02$ ) that remained undetected by the pseudomarker method. This suggests that although the risk of false positive results may increase with the multiple tests used, the comparison of the methods would be important for not losing the true ones. Indeed, the region 4q26 has shown linkage to CD in the UK population (King et al., 2000&2001), and 15q11 was the most promising candidate gene region in our independent genome scan in CD families from a Finnish sub-isolate (Woolley et al., 2002). As we could not find linkage in the candidate region 15q26 (II), located at least 80 cM telomeric from 15q11, it will also be interesting to see whether there is two independent susceptibility factors, if any, present on chromosome 15q, or whether the difference in the location just results from the poor resolution of linkage approaches for a single susceptibility locus.

Conclusively, the numerous genome-wide and candidate gene analyses on celiac disease have not yet revealed definitive evidence for any HLA-unlinked locus. The most promising regions showing linkage in many independent studies include the chromosomes 2q, 5q and 11q which were also supported by the studies in this thesis. These regions harbor several pathogenetically relevant candidate genes, but any disease-associated polymorphisms have not yet been identified. Linkage evidence in all candidate regions varies from weak to moderate, compared to the highly significant linkage scores observed in the HLA region. The HLA-unlinked genes involved in celiac disease are therefore likely to play only a minor and obviously a more complex role in disease susceptibility.

## THE ROLE OF HLA - DOSE-EFFECT AND SEX-MODULATED RISK DUE TO DQ2

The role of HLA-DQ in CD susceptibility was evident in the present family material. Linkage in the HLA region on 6p23.1 was highly significant (LOD=5.96) in the 60 families studied genome-wide, and among all members in 98 families an extremely high LOD score of 19.6 was found at DQB1 locus. Highly significant linkage was also found in all subgroups of families divided according to the disease phenotypes and sex (I). All but one affected individual carried either the DQB1\*02 (DQ2) or DQB1\*0302 (DQ8) susceptibility alleles. The CD prevalence among all DQB1\*02 positive family members was as high as 65%. However, since the families were selected to have multiple patients, this can not be compared with the overall population risk of

5% among DQ2 positive individuals, if the 1% prevalence of CD (Kolho et al., 1998) and 20% frequency of DQ2 (Partanen and Westman, 1997) are assumed. Although one copy of the genes encoding the DQ2 molecule is enough for the disease onset, the dose dependence of CD susceptibility due to the number of carried HLA-DQ2 molecules has been suggested in several studies (Ploski et al., 1993; Congia et al., 1994). The difference was also seen in our family sample: 79% of all DQB1\*02 homozygotes were affected with CD, compared to the 60% prevalence among heterozygotes. Due to the non-independence of family members, the statistical significance of this difference can not, however, be determined.

Females have a higher risk of CD than males, with a sex ratio of 2:1 (Logan, 1992), which is a typical finding in many autoimmune diseases (Whitacre et al., 1999). The reason for this is unknown, but hormonal differences and immunological effects of pregnancies have been suggested. In the case of celiac disease, which often manifests with only mild or atypical symptoms, the sex difference can also result from a lower threshold for females in going to the clinic. This is interestingly supported by the fact that DH, which manifests with a strongly itching and clearly visible rash, shows rather male than female overrepresentation among the patients (Reunala, 1996). In this thesis, the sex ratio among the patients of 102 families studied in the study I was 2.2:1, and importantly, among all HLA-DQ2 positive family members, the risk of CD was significantly higher in females than males ( $p=0.0001$ ). The high overall CD prevalence in DQ2 positive individuals in our families is definitely biased due to the selection of multiply affected families and the genetic non-independence of family members, but these should not bias the comparison between the sexes. The result therefore suggests that the DQ2 alone is a more stronger risk factor for females. Therefore, the additional genetic risk factors could play a stronger role in the disease onset of males. This was indeed supported by our finding that linkage at three non-HLA candidate loci was significant only in families with at least one male patient with CD (I). Although the existence of true linkage heterogeneity could not be equivocally proven due to small sample sizes, this may support a gain of power by studying families with affected males.

## **FAILURES IN REPLICATING LINKAGE - REASONS?**

*Sample size.* The reasons for the discrepancies between several linkage studies on CD may lie in true heterogeneity for the disease susceptibility loci between the studied populations, or the higher rate of false positive and negative linkage results which is a known problem in studies of complex diseases. Under locus heterogeneity and other confounding factors in these studies, the statistically powerful sample size required for the detection of genes with a small effect on disease can be very large, often unrealistic for collection or study. One critical problem in most of the linkage studies performed for CD, including ours, is the lack of power estimations for the used family samples. However, strong linkage to the HLA region was

## DISCUSSION

observed in most of these studies. This clearly indicates that the power is high enough to detect a locus with a strong risk effect. Power calculations for sample size needed are specific to the risk allele prevalence and the strength of its risk effect - assumptions which are not easy to assess for a polygenic disease. As only suggestive linkage scores are usually found in most genome scans of any multifactorial disease, more important than reaching the significant linkage level in an individual scan would be the support of the results by independent studies.

*Sample type and risk effects of the genes.* The detection of loci with different allele frequencies and magnitudes of risk effect can also depend on the structures of families included in the analysis. Linkage analysis in large pedigrees are likely to be optimal for rare risk alleles with relatively strong effect, since the probability of such an allele segregating in a large family is increased, whereas in large collections of affected sib-pairs the heterogeneity for a rare locus might be too high for detection of linkage. Conversely, risk alleles with a high frequency but relatively weak effect are more likely to occur as homozygous in parents and hence, their detection may be difficult in an analysis including a low number of large pedigrees. The larger samples of affected sib-pairs which are also usually easier to collect, are more robust to this effect. Indeed, the success in replication of linkage evidence among the studies on celiac disease seems to correlate with the type of the family material used (Naluai et al., 2001). Positive evidence for chromosomes 11p and 15q were observed only in those studies using extended, multiply affected families, whereas in studies with affected sib-pairs, evidence for linkage in 5q and 11q was accumulated. Indeed, the high population frequency of the known susceptibility HLA alleles can also explain the failure to detect linkage at HLA locus in the genome-wide study with extended pedigrees by King et al (2000), although all patients in these families carried either HLA-DQ2 or DQ8. This was also demonstrated in our recent genome-scan for nine distantly related families from the North-Eastern Finland, including high number of CD patients: in spite of interesting results in a few non-HLA regions, no linkage to HLA could be detected since the number of the segregating HLA-DQ2 haplotypes in these families was high (Woolley et al., 2002).

*Comparison of the studies – meta-analysis.* One current attempt to compare and combine the results of genome scans is meta-analysis. This method can increase the power to detect loci with a weak effect, for which the linkage scores in individual scans might have remained under the cut-off level. However, the level of locus heterogeneity among study samples from different populations will be higher. The effect of this to the meta-analysis remains to be seen; the risk is always that evidence for regions observed in only some populations is going to disappear in the combined material. A meta-analysis on Finnish (V), Italian, Swedish and British genome-wide results is ongoing, and will hopefully strengthen the evidence for at least some of the suggested regions and probably explain some of the discrepancies between the studies. Indeed, encouraging evidence in favor of the chromosomes 2q, 5q and 11q has been obtained (unpublished).

*Minimization of heterogeneity – subgroups within the trait.* To overcome the problem of heterogeneity within a study material, at least two options for the sampling strategies exist. The first can be applied to any population under study. Linkage can be tested separately in subgroups of families divided according to any phenotypic feature of the disease (such as DH and non-DH among celiac patients), or some already established genetic risk factor (such as HLA-DQ in CD). This can be assumed to result from a more homogeneous genetic background among the selected patients. The problem in this method is the decrease in the sample size, affecting the power. Our results in publication I showed some preliminary evidence for genetic differences between two manifestations of celiac disease, as well as between the sex of patients. However, the determination of the disease phenotypes in celiac disease can be problematic, since the gluten-free diet, started due to the first disease manifestation will effectively prevent the onset of other forms of the disease. In addition, as the disease onset can occur at any age, the follow-up of these family members in respect of genetic analyses is difficult. This also raises the question of whether linkage methods based only on affected individuals should be used for celiac disease. This problem of phenotypic determination can probably be partly overtaken in future if any non-HLA gene involved in CD will be identified. The stratification of the patients or families due to that genetic factor is very likely to help in the search for other remaining susceptibility loci. The stratification due to the HLA susceptibility genes is already possible, but as only a few families had patients with the DQ8 risk allele, the division of the families into DQ2 positive and DQ2 negative groups was not meaningful in our family sample. Instead, the classification of the patients to 'high' and 'very high' risk groups due to the number of HLA risk alleles carried by them would be interesting, especially as our own results supported the dose-dependent effect of CD susceptibility due to HLA-DQ2 (V). The relative role of HLA-unlinked susceptibility genes could therefore be stronger in patients carrying only one copy of DQ2. Indeed, a similar classification of the family sample strengthened the linkage evidence at the CTLA4 gene region in the genome-wide scan by Naluai et al (2001).

*Minimization of heterogeneity – population choice.* Since the level of heterogeneity for genetic and environmental factors is likely to be higher in more mixed populations, another option to minimize it is the use of isolated populations. They can be assumed to be more homogeneous for the susceptibility genes and the founder effect present in these populations is also useful in linkage disequilibrium based methods for positional cloning of these genes (Wright et al., 1999). The Finnish population is a good example of a founder population, and has shown its power in mapping studies of rare recessive monogenic diseases overrepresented in Finland (Peltonen et al., 1999). For many relatively common single gene disorders with dominant inheritance, the strong founder effect is visible as well. These diseases often show distinct founder mutations enriched to regional subpopulations of Finland (Kere, 2001). A good example is familial hypercholesterolemia (FH), caused by mutations in the low-density lipoprotein (LDL) receptor gene which leads to elevated plasma LDL levels and prematurely developed

## DISCUSSION

atherosclerosis. Approximately 1 in 500 individuals are heterozygous carriers of FH mutations in the general population and over 600 different mutations have been described worldwide (Ose, 1999). In Finland, four major mutations account for 75% of all FH patients and show high differences in regional frequencies (Vuorio et al., 2001), still being rare in other Nordic countries (Lind et al., 1998). Studies on FH have also demonstrated a wide phenotypic diversity in this monogenic disease which was earlier considered very monotonous. In fact, it has been suggested that most of the simple mendelian disorders are likely to have variation in the clinical phenotype due to the other genetic background and environmental influences, thus actually sharing many analytical problems with the multifactorial diseases (Dipple and McCabe, 2000).

The advantages of the Finnish population for studies on complex diseases still remain to be established (Kere, 2001). Although several genome-wide studies have been performed, the success of them can not be compared to studies in other populations as long as the actual susceptibility alleles in these linked regions remain unidentified. Even if the total number of involved risk loci could be assumed to be lower in Finland, the observed geographic differences in the prevalence of some of these diseases do not directly indicate the enrichment of single susceptibility alleles, which is the case for monogenic diseases with high penetrance. Furthermore, the level of linkage disequilibrium in the general Finnish population may not be significantly higher than in other populations, as opposed to the rare founder mutations of Finnish disease heritage (Eaves et al., 2000). In our genome scan among Finnish unrelated CD families (V), no single non-HLA region was revealed with significant linkage, which would have been indicative for increased homogeneity of Finnish celiac families for some of the risk genes. In fact, the scan result was very similar to most of the other scans of autoimmune diseases, showing multiple positive peaks with weak linkage evidence, and a strong linkage only in the HLA region. Therefore the level of homogeneity for celiac disease susceptibility genes in the Finnish population might not differ significantly from more mixed populations. One further way to reduce heterogeneity could be the use of sub-isolates within the Finnish population. In addition to the assumed genetic similarity, the patients of these are likely to share more identical environment, which is an important factor in susceptibility to complex disorders. Indeed, our recent genome-wide analysis for nine distantly related families with CD, from the North-Eastern Finland, showed significant linkage in one novel region and supported some of the findings in publication V (Woolley et al., 2002). Encouraging linkage results from Finnish sub-isolates have been reported also for hypertension (Perola et al., 2000) and asthma (Laitinen et al., 2001).

## CELIAC DISEASE AS A COMPLEX GENETIC DISORDER

As discussed above, the common problems of heterogeneity and other confounding factors also complicate the genetic studies of celiac disease. However, among other autoimmune type of diseases, celiac disease features several advantages. The major environmental trigger, dietary gluten, is known, which is not the case for other autoimmune type of diseases. Also, this agent is encountered by practically all individuals. The HLA association of CD is one of the strongest among these diseases, and the functional role of the HLA risk molecules in disease pathogenesis has been resolved. For the *in situ* studies, access to the affected organ, the small intestine, is simple and can be performed safely. Since the diagnosis of CD and many other intestinal diseases is based on small bowel biopsy, the collection of both CD and non-CD tissue samples is relatively easy.

One drawback in studies on celiac disease is the lack of an appropriate animal model, which are available for many other autoimmune diseases, e.g. the non-obese diabetic (NOD) mice as a model for type I diabetes (Wicker et al., 1995), experimental autoimmune encephalomyelitis (EAE) and neuritis (EAN) for demyelinating disorders of the nervous system (Gold et al., 2000) and collagen-induced arthritis (CIA) for human rheumatoid arthritis (Luross and Williams, 2001). In addition to their usefulness in immunopathological studies of these diseases, the linkage mapping of risk genes in the inbred animal strains also has several advantages over the human studies (Risch, 2000). The level of heterogeneity can be minimized, rare disease alleles can be fixed in the strain by positive selection in many generations, and information in linkage analysis is very high when fully heterozygous animals from the mating of two inbred strains are used as parents. In addition, the environment of animals is controllable and can be standardized, and the offspring of the same litter are at an equal age and share an identical environment. These factors will minimize the phenotypic variance of the disease due to non-genetic factors. The linkage analyses for locating the genomic areas carrying risk factors have shown to be powerful with these animal models and multiple regions homologous to the linked areas in human studies have been found (Wandstrat and Wakeland, 2001). However, the positional cloning of the actual risk genes has still remained a challenge. Alarmingly, these animal studies have revealed an involvement of the modifier or suppressive genetic loci, which are fully protective over the risk alleles of autoimmunity (Morel et al., 1999). Assumingly these kind of genetic interactions are likely to underlie in human diseases as well, causing further complexity in the search for the susceptibility genes. In spite of several advantages of animal models, it should be kept in mind that the similar disease phenotypes of animals and humans might not result from a fully identical genetic and pathogenetic background. Indeed, the existence of Irish Setter dogs suffering from naturally occurring gluten sensitive enteropathy (GSE) which resembles celiac disease and shows histologically similar reversible intestinal injury triggered by gluten, has encouraged the researchers to investigate these dogs as model

## *DISCUSSION*

animals for CD (Batt et al., 1987). However, in large dog pedigrees highly affected by GSE, no role of MHC genes could be detected (Polvi et al., 1998). Contradictory to celiac disease, where the HLA molecules play a crucial role in disease susceptibility and pathogenesis, this indicates that the immunopathogenetic mechanisms might differ significantly between the dog and human diseases.

In the future, the genetic studies on celiac disease will hopefully benefit from both other fields of studies on CD and from recent advances in modern genetics. For the past two years, a large European consortium of 17 study groups including ours, has co-operated in order to reveal new and reliable data on celiac disease epidemiology, pathogenetic mechanisms and genetic susceptibility factors in a large, combined patient material. For genetic analyses, over 800 CD families have been collected by six groups. The studies on different fields of celiac disease are also likely to support each other. The linkage evidence in a given region always needs further studies to characterize the primary gene and allele involved with the disease, and subsequently to prove the functional effect of this genetic factor on a given phenotype. If this is known for any of the HLA-unlinked susceptibility genes in future, this information will naturally provide help for the epidemiological and pathogenetic studies on CD. Novel epidemiological data such as the disease prevalence among the carriers or concordance for the siblings stratified for the given genetic factor will be available to collect. This can then help to estimate the role of this factor in the total genetic background, as well as possible inheritance models or epistatic gene-gene interactions involved with it. Together with the recent analytical advances for e.g. simultaneous tests for multiple loci, these models will hopefully increase the power of linkage studies in the search for the remaining susceptibility genes. Studies on the pathogenetic mechanisms involved in celiac disease will naturally also gain advantage from the identified genetic factors and could further suggest new functional candidate genes to test for genetic linkage and association in family samples. With the new high-throughput microarray approaches to study gene expression differences between affected and unaffected tissue samples, the search for novel candidates involved in the pathogenesis or genetic susceptibility of the disease may be feasible. The completion of human genome sequencing will offer a large pattern of novel polymorphisms, candidate genes and accurate marker maps for genetic studies. Together with novel cost- and labor-effective, high-throughput methods for genotyping, and with statistical improvements such as data-mining approaches for this strongly expanding genotypic data, the challenge of detecting even the minor susceptibility genes of complex diseases may turn out to be a realistic task.

## CONCLUDING REMARKS

Although the evidence for strong genetic background of celiac disease is clear, only the major susceptibility genes at HLA-DQ have been established so far. HLA-DQ2 or DQ8 molecules are necessary but not sufficient alone to the disease onset. In this thesis, we searched for HLA-unlinked susceptibility factors of celiac disease. We concentrated on previously suggested candidate gene regions and found supporting evidence for possible susceptibility factors present on chromosomes 5q and 11q (I), while no clear linkage was observed at 15q26 (II). We also showed the first lines of evidence for genetic linkage between celiac disease and the CTLA4/CD28/ICOS gene region on 2q33 (III). However, the identification of the primary disease-associated variation in this gene region is complicated, which was demonstrated by our results showing a high level of linkage disequilibrium within the CTLA4 gene (IV). Our genome-wide search for the novel susceptibility loci involved in celiac disease revealed evidence for six non-HLA regions which partially supported the findings of other studies (V). HLA still seems to be the major risk locus and the other genetic factors may play a minor or a more complex role in CD. The linkage results in this thesis and from other populations create together a basis for the future search for genuine susceptibility genes at least on chromosomes 2q, 5q and 11q. No clear advantage of concentrating on the Finnish population was seen from the genome-wide results, suggesting that a remarkable level of heterogeneity in complex diseases may occur even in a founder population. The linkage heterogeneity for the candidate regions was suggested between the disease phenotypes and the sex (I). This, together with the discrepancies between the genetic studies on celiac disease and with the observed sex- and dose-dependent role of HLA risk factors (I, V), points to the complexity of the underlying genetic factors involved in celiac which is likely to complicate the genetic analyses of this multifactorial disease.

## ACKNOWLEDGEMENTS

This study was carried out during the years 1998-2002 at the Department of Tissue Typing of the Finnish Red Cross Blood Transfusion Service in Helsinki. I want to thank sincerely the former and present Directors of the Institute, Professor Juhani Leikola and Docent Jukka Rautonen; the Director of the former Division of Laboratory Services and the present Division of Blood and Blood Components, Docent Tom Krusius; the Director of the Division of Stem Cell and Transplantation Services, Docent Jarmo Laine, and the previous and present Heads of the Tissue Typing Laboratory, Professor Saija Koskimies and Docent Jukka Partanen, for providing the excellent research facilities. Tom and Jukka P. are especially respected for their honest care and warm attitude to me as a student.

I was privileged to have Docent Jukka Partanen as my supervisor. What else could I say but thank you for everything! Especially I appreciate your humanity, great sense of humor, the friendly way of encouraging to use my own brains and your trust on me. Our various discussions more or less around science have been the main source of my motivation on this work.

This work would have been impossible without the excellent collaboration with the clinicians in Tampere University Hospital and Institute of Medical Technology in University of Tampere. I express my sincere thanks to Professor Markku Mäki, Dr. Kirsi Mustalahti and Docent Pekka Collin for providing us with the patient material. The whole Celiac Disease Study Group is thanked for their exceptionally cheerful attitude to everything. During any scientific discussion, I have never laughed as much as every time visiting Tampere. I am grateful to you, Markku, also for guiding me to this project in the first place, and for your admirable enthusiasm on celiac disease research which feels always inspiring.

I want to express my deepest gratitude to our important collaborators in Columbia University in New York - Jianjun Liu, Suh-Hang Juo, Joe Terwilliger, Xiaomei Tong, Adina Grunn, Miguel Brito, Peter Green and Conrad Gilliam – for the huge effort on the genome-wide study and for the fruitful and educative co-operation. The European collaborators, you all nice people from France, Italy, Norway, Sweden and UK, are also thanked for showing that an effective but still so pleasant and friendly way of team work is possible.

I am grateful to Docents Erkki Savilahti and Pentti Tienari for the critical revision and very valuable comments on this thesis. The Division of Genetics in University of Helsinki, and Professor Hannu Saarilahti are appreciated for the uncomplicated manner to handle with post-graduate studies.

All the young researcher colleagues and dear friends in the Blood Transfusion Service – Anne, Antti, Elisa, Iiris, Kaisa, Kati, Katri, Kristiina, 2xLaura, Leena, Lotta, Meri, Mervi, Miia, Mikko, Niina, Nina, Paula, Pekka, Rosa, 2xSatu, Taina and Tuuli - are thanked for the almost-everlasting optimism and the numerous cheerful moments during the long days in the lab and after work. The special ones of you have really changed the meaning of *työkaveri*. A big hug for your friendship! Katri, the multi-talented one, is thanked for being also my target in the fencing area and for helping me to forget the work with the various other sporting activities. The other members of our badminton team, Kati and Niina, as well as our coach Eero Hämäläinen are appreciated for sharing the many hours of fun, sweat and fat-burning. I will miss so much the Wednesday sauna with the relaxing discussions.

I have been happy to work in the Tissue Typing Lab. Docent Maisa Lokki, Doctors Irma Matinlauri and Kimmo Talvensaari, as well as the whole Tissue Typing staff full of fantastic people are warmly thanked for all kinds of help, the friendly attitude and genuine interest on my studies. Marjaana Mustonen is especially acknowledged for an enormous help in sample handling and HLA typing of the patient material of this thesis. ALL of you, thanks for having always been so nice to me.

## ACKNOWLEDGEMENTS

Docent Pertti Sistonen, Docent Pekka Uimari, Sonja Lumme and Antti Penttilä are thanked for the valuable statistical and computing help for this thesis. With Sonja as well as Katri, my soul-mates in clumsiness, I have really enjoyed our sane capability to laugh at ourselves.

Huge thanks to James Woolley for revising the language of this thesis. I am very grateful to Miikka Haimila for the enormous help and patience during the lay-out process. Marja-Leena Hyvönen and Maija Ekholm are thanked for the excellent library services.

To my parents Eeva-Maija and Markku I owe big thanks for always believing in me and supporting me in all possible ways. You have taught me the most important lesson to just trying to do the best which should always be enough for everything. You and my dear brother Jukka, as well as other relatives are also thanked for the honest interest in the field that I'm working in. I feel grateful to all my friends from the old times and from the student years in Jyväskylä for providing me with the great memories to carry with delight. Jaana, a big hug for the twenty-years friendship – our special, strange but so hilarious, sense of humor will last forever. I also feel lucky to have three wonderful god-children Jenni-Mari, Rasmus and Emma. The joy of living in your face is the best way to remind me of the other amazing aspects of life.

My energetic flatmate Eliisa is thanked for so many evenings full of laugh and her always-ready-to-go attitude in a case of an urgent need of a drink in the downstairs 'living room'. I am grateful to her and the marvelous boys of the Restaurant Delicatessen for the huge help in organizing the party.

Warm thanks to Eerikki for the many shared years and for keeping me going to work and for trying to get me out of there in a decent time. Thanks for the love and the friendship I appreciate a lot.

I am indebted to all voluntary Finnish CD families who have made these studies possible in the first place. The copyright holders are acknowledged for their permission to reproduce the original articles as a part of this publication.

This study was financially supported by Emil Aaltonen Foundation, Sigrid Juselius Foundation, Maud Kuistila Memorial Foundation, Medical Research Funds of Finnish Red Cross Blood Transfusion Service and Tampere University Hospital, University of Helsinki, Medical Research Council of the Academy of Finland, and the Commission of the European Communities, specific RTD programme "Quality of Life and Management of Living Resources", QLRT-1999-00037, "Evaluation of the prevalence of the coeliac disease and its genetic components in the European population". It does not necessarily reflect its views and in no way anticipates the Commission's future policy in this area.

Finally, I congratulate myself for just trying to do my best. Some might say that during these years of too much work I have not learned to live. I still believe that I have learned at least something about *life* - and choices.

Helsinki, April 23<sup>th</sup> 2002



Päivi Holopainen

## REFERENCES

- Aguado B, Milner CM, Campbell RD (1996) Genes of the MHC class III region and the functions of the proteins they encode. In: Browning M, McMichael A (eds) HLA and MHC: genes, molecules and function. BIOS Scientific Publishers Ltd, Oxford, pp 39-75
- Ali-Varpula N, Holopainen P, Bourgain C, Mustalahti K, Collin P, Mäki M, Partanen J (2002) CD80 (B7-1) and CD86 (B7-2) genes and genetic susceptibility to coeliac disease. *Eur J Immunogenet* (in press)
- Anand BS, Piris J, and Truelove SC (1978) The role of various cereals in coeliac disease. *Q J Med* 47: 101-110
- Anderson RP, Degano P, Godkin AJ, Jewell DP, and Hill AV (2000) In vivo antigen challenge in celiac disease identifies a single transglutaminase-modified peptide as the dominant A-gliadin T-cell epitope. *Nat Med* 6: 337-342
- Arai T, Michalski JP, McCombs CC, Elston RC, McCarthy CF, and Stevens FM (1995) T cell receptor gamma gene polymorphisms and class II human lymphocyte antigen genotypes in patients with celiac disease from the west of Ireland. *Am J Med Sci* 309: 171-178
- Arentz-Hansen H, Körner R, Molberg O, Quarsten H, Vader W, Kooy YM, Lundin KE, Koning F, Roepstorff P, Sollid LM, and McAdam SN (2000) The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. *J Exp Med* 191: 603-612
- Auricchio S, Follo D, de Ritis G, Giunta A, Marzorati D, Prampolini L, Ansaldi N, Levi P, Dall'Olio D, and Bossi A (1983) Does breast feeding protect against the development of clinical symptoms of celiac disease in children? *J Pediatr Gastroenterol Nutr* 2: 428-433
- Bardella MT, Fredella C, Prampolini L, Marino R, Conte D, and Giunta AM (2000) Gluten sensitivity in monozygous twins: a long-term follow-up of five pairs. *Am J Gastroenterol* 95: 1503-1505
- Batt RM, McLean L, and Carter MW (1987) Sequential morphologic and biochemical studies of naturally occurring wheat-sensitive enteropathy in Irish setter dogs. *Dig Dis Sci* 32: 184-194
- Becker KG, Simon RM, Bailey-Wilson JE, Freidlin B, Biddison WE, McFarland HF, and Trent JM (1998) Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. *Proc Natl Acad Sci USA* 95: 9979-9984
- Bell GI, Selby MJ, and Rutter WJ (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature* 295: 31-35
- Bevan S, Popat S, Braegger CP, Busch A, O'Donoghue D, Falth-Magnusson K, Ferguson A, Godkin A, Hogberg L, Holmes G, Hosie KB, Howdle PD, Jenkins H, Jewell D, Johnston S, Kennedy NP, Kerr G, Kumar P, Logan RF, Love AH, Marsh M, Mulder CJ, Sjoberg K, Stenhammer L, and Houlston RS (1999) Contribution of the MHC region to the familial risk of coeliac disease. *Journal of Medical Genetics* 36: 687-690
- Bonamico M, Mariani P, Danesi HM, Crisogianni M, Failla P, Gemme G, Quartino AR, Giannotti A, Castro M, Balli F, Lecora M, Andria G, Guariso G, Gabrielli O, Catassi C, Lazzari R, Balocco

## REFERENCES

- NA, De Virgiliis S, Culasso F, and Romano C (2001) Prevalence and clinical picture of celiac disease in Italian Down Syndrome patients: a multicenter study. *J Pediatr Gastroenterol Nutr* 33: 139-143
- Bouguerra F, Dugoujon JM, Babron MC, Greco L, Khaldi F, Debbabi A, Bennaceur B, and Clerget-Darpoux F (1999) Susceptibility to coeliac disease in Tunisian children and GM immunoglobulin allotypes. *Eur J Immunogenet* 26: 293-297
- Brett PM, Yiannakou JY, Morris M-A, Bronson SR, Mathew C, Curtis D, and Ciclitira PJ (1998) A pedigree-based linkage study of coeliac disease: failure to replicate previous positive findings. *Ann Hum Genet* 62: 25-32
- Broman KW, and Weber JL (1998) Estimation of pairwise relationships in the presence of genotyping errors. *Am J Hum Genet* 63: 1563-1564
- Bugeon L, and Dallman MJ (2000) Costimulation of T cells. *Am J Respir Crit Care Med* 162: S164-S168
- Carbonara AO, DeMarchi M, van Loghem E, and Ansaldi N (1983) Gm markers in celiac disease. *Hum Immunol* 6: 91-95
- Cardon LR, and Bell JI (2001) Association study designs for complex diseases. *Nat Rev Genet* 2: 91-99
- Caruso C, Candore G, Lio D, Modica MA, Cataldo F, Maltese I, Marino V, and Albeggiani A (1991a) Immunoglobulin heavy-chain allotypes play a role in the clinical history of celiac disease. *Exp Clin Immunogenet* 8: 85-87
- Caruso C, Candore G, Modica MA, Lio D, Marino V, Maltese I, Cataldo F, and Albeggiani A (1991b) Immunoglobulin heavy chain allotypes in a sample of Sicilian patients with celiac disease. *Exp Clin Immunogenet* 8: 1-5
- Cataldo F, Marino V, Ventura A, Bottaro G, and Corazza GR (1998) Prevalence and clinical features of selective immunoglobulin A deficiency in coeliac disease: an Italian multicentre study. Italian Society of Paediatric Gastroenterology and Hepatology (SIGEP) and "Club del Tenue" Working Groups on Coeliac Disease. *Gut* 42: 362-365
- Catassi C, Ratsch IM, Fabiani E, Rossini M, Bordicchia F, Candela F, Coppa GV, and Giorgi PL (1994) Coeliac disease in the year 2000: exploring the iceberg. *Lancet* 343: 200-203
- Chotai J (1984) On the lod score method in linkage analysis. *Ann Hum Genet* 48: 359-378
- Clot F, Babron MC, Percopo S, Giordano M, Bouguerra F, Clerget-Darpoux F, Greco L, Serre JL, and Fulchignoni-Lataud MC (2000) Study of two ectopeptidases in the susceptibility to celiac disease: two newly identified polymorphisms of dipeptidylpeptidase IV. *Journal of Pediatric Gastroenterology & Nutrition* 30: 464-466
- Clot F, Fulchignoni-Lataud MC, Renoux C, Percopo S, Bouguerra F, Babron MC, Djilali-Saiah I, Caillat-Zucman S, Clerget-Darpoux F, Greco L, and Serre JL (1999a) Linkage and association study of the CTLA-4 region in coeliac disease for Italian and Tunisian populations. *Tissue Antigens* 54: 527-530
- Clot F, Gianfrani C, Babron MC, Bouguerra F, Southwood S, Kagnoff MF, Troncone R, Percopo S, Eliaou JF, Clerget-Darpoux F, Sette A, and Greco L (1999b) HLA-DR53 molecules are associ-

## REFERENCES

- ated with susceptibility to celiac disease and selectively bind gliadin-derived peptides. *Immunogenetics* 49: 800-807
- Collin P, Kaukinen K, and Mäki M (1999) Clinical features of celiac disease today. *Dig Dis* 17: 100-106
- Collin P, Mäki M, Keyriläinen O, Hällström O, Reunala T, and Pasternack A (1992) Selective IgA deficiency and coeliac disease. *Scand J Gastroenterol* 27: 367-371
- Collin P, Reunala T, Pukkala E, Laippala P, Keyriläinen O, and Pasternack A (1994) Coeliac disease - associated disorders and survival. *Gut* 35: 1215-1218
- Collin P, Salmi J, Hällström O, Oksa H, Oksala H, Mäki M, and Reunala T (1989) High frequency of coeliac disease in adult patients with type-I diabetes. *Scand J Gastroenterol* 24: 81-84
- Collin P, Salmi J, Hällström O, Reunala T, and Pasternack A (1994) Autoimmune thyroid disorders and coeliac disease. *Eur J Endocrinol* 130: 137-140
- Congia M, Cucca F, Frau F, Lampis R, Melis L, Clemente MG, Cao A, and De Virgiliis S (1994) A gene dosage effect of the DQA1\*0501/DQB1\*0201 allelic combination influences the clinical heterogeneity of celiac disease. *Hum Immunol* 40: 138-142
- Cooper BT, Holmes GK, and Cooke WT (1978) Coeliac disease and immunological disorders. *Br Med J* 1: 537-539
- Cottingham Jr RW, Idury RM, and Schäffer AA (1993) Faster sequential genetic linkage computations. *Am J Hum Genet* 53: 252-263
- Cronin CC, Feighery A, Ferriss JB, Liddy C, Shanahan F, and Feighery C (1997) High prevalence of celiac disease among patients with insulin-dependent (type I) diabetes mellitus. *Am J Gastroenterol* 92: 2210-2212
- Daum S, Bauer U, Foss HD, Schuppan D, Stein H, Riecken EO, and Ullrich R (1999) Increased expression of mRNA for matrix metalloproteinases-1 and -3 and tissue inhibitor of metalloproteinases-1 in intestinal biopsy specimens from patients with coeliac disease. *Gut* 44: 17-25
- Deichmann K, Heinzmann A, Bruggenolte E, Forster J, and Kuehr J (1996) An Mse I RFLP in the human CTLA4 promotor. *Biochem Biophys Res Commun* 225: 817-818
- DeMarchi M, Borelli I, Olivetti O, Richiardi P, Wright P, Ansaldi N, Barbera C, and Santini B (1979) Two HLA-D and DR alleles are associated with coeliac disease. *Tissue Antigens* 14: 309-316
- Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, and Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380: 152-154
- Dicke, W. K. Coeliakie: een onderzoek naar de nadelige invloed van sommige graansorten op de lijder aan coeliakie. 1950. University of Utrecht.
- Dieterich W, Ehnis T, Bauer M, Donner B, Volta U, Riecken EO, and Schuppan D (1997) Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nat Med* 3: 797-801

## REFERENCES

- Dieterich W, Storch WB, and Schuppan D (2000) Serum antibodies in celiac disease. *Clin Lab* 46: 361-364
- Dipple KM, and McCabe ER (2000) Phenotypes of patients with "simple" Mendelian disorders are complex traits: thresholds, modifiers, and systems dynamics. *Am J Hum Genet* 66: 1729-1735
- Djilali-Saiah I, Schmitz J, Harfouch-Hammoud E, Mougenot J-F, Bach J-F, and Caillat-Zucman S (1998) CTLA-4 gene polymorphism is associated with predisposition to coeliac disease. *Gut* 43: 187-189
- Donis-Keller H, Green P, Helms C, Cartinour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, and Lander ES (1987) A genetic linkage map of the human genome. *Cell* 51: 319-337
- Duffy DL (1997) Sib-pair: a program for non-parametric linkage/association analysis. *Am J Hum Genet* 61(suppl): A197
- Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, and Todd JA (2000) The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nat Genet* 25: 320-323
- Ek J, Albrechtsen D, Solheim BG, and Thorsby E (1978) Strong association between the HLA-Dw3-related B cell alloantigen-DRw3 and coeliac disease. *Scand J Gastroenterol* 13: 229-233
- Ellis A (1981) Coeliac disease: previous family studies. In: McConnel RB (ed) *The genetics of coeliac disease*. MTB, Lancaster, pp 197-200
- Falchuc ZM, Rogentine GN, and Strober W (1972) Predominance of histocompatibility antigen HL-A8 in patients with gluten-sensitive enteropathy. *J Clin Invest* 51: 1602-1605
- Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, Dormishian F, Domingo RJ, Ellis MC, Fullan A, Hinton LM, Jones NL, Kimmel BE, Kronmal GS, Lauer P, Lee VK, Loeb DB, Mapa FA, McClelland E, Meyer NC, Mintier GA, Moeller N, Moore T, Morikang E, and Wolff RK (1996) A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* 13: 399-408
- Fedrick JA, Pandey JP, Verkasalo M, Teppo AM, and Fudenberg HH (1985) Immunoglobulin allotypes and the immune response to wheat gliadin in a Finnish population with celiac disease. *Exp Clin Immunogenet* 2: 185-190
- Ferguson A, Arranz E, and O'Mahony S (1993) Clinical and pathological spectrum of coeliac disease - active, silent, latent, potential. *Gut* 34: 150-151
- Fernández-Arquero M, Figueredo MA, Maluenda C, and de la Concha EG (1995) HLA-Linked genes acting as additive susceptibility factors in celiac disease. *Hum Immunol* 42: 295-300
- Field LL, Tobias R, and Magnus T (1994) A locus on chromosome 15q26 (IDDM3) produces susceptibility to insulin-dependent diabetes mellitus. *Nat Genet* 8: 189-194
- Fry L, Seah PP, Riches DJ, and Hoffbrand AV (1973) Clearance of skin lesions in dermatitis herpetiformis after gluten withdrawal. *Lancet* 1: 288-291
- Gee S (1888) On the coeliac affection. *St Bart Hospital Rep* 24: 17-20

## REFERENCES

- Giannotti A, Tiberio G, Castro M, Virgili F, Colistro F, Ferretti F, Digilio MC, Gambarara M, and Dallapiccola B (2001) Coeliac disease in Williams syndrome. *J Med Genet* 38: 767-768
- Gold R, Hartung HP, and Toyka KV (2000) Animal models for autoimmune demyelinating disorders of the nervous system. *Mol Med Today* 6: 88-91
- Greco L, Babron MC, Corazza GR, Percopo S, Sica R, Clot F, Fulchignoni-Lataud MC, Zavattari P, Momigliano-Richiardi P, Casari G, Gasparini P, Tosi R, Mantovani V, De Virgiliis S, Iacono G, D'Alfonso A, Selinger-Leneman H, Lemainque A, Serre JL, and Clerget-Darpoux F (2001) Existence of a genetic risk factor on chromosome 5q in Italian Coeliac Disease families. *Ann Hum Genet* 65: 35-41
- Greco L, Corazza G, Babron MC, Clot F, Fulchignoni-Lataud M-C, Percopo S, Zavattari P, Bouguerra F, Dib C, Tosi R, Troncione R, Ventura A, Mantovani W, Magazzu G, Gatti R, Lazzari R, Giunta A, Perri F, Iacono G, Cardi E, De Virgiliis S, Cataldo F, Musumeci S, Ferrari R, Balli F, Bardella MT, Volta U, Catassi C, Torre G, Eliaou J-F, Serre J-L, and Clerget-Darpoux F (1998) Genome search in celiac disease. *Am J Hum Genet* 62: 669-675
- Greco L, Mäki M, Di Donato F, Visakorpi JK (1992) Epidemiology of coeliac disease in Europe and the Mediterranean area. In: Auricchio S, Visakorpi JK (eds) *Common food intolerances 1: Epidemiology of coeliac disease*. Dynamic Nutrition Research, Basel, pp 25-44
- Greco L, Romino R, Coto I, Di Cosmo N, Percopo S, Maglio M, Paparo F, Gasperi V, Limongelli MG, Cotichini R, D'Agate C, Tinto N, Sacchetti L, Tosi R, Stazi MA (2002) The first large population based twin study of coeliac disease. *Gut* 50: 624-628
- Grillo R, Petronzelli F, Mora B, Bonamico M, and Mazzilli MC (2000) Search for coeliac disease susceptibility loci on 7q11.23 candidate region: absence of association with the ELN17 microsatellite marker. *Hum Hered* 50: 180-183
- Göring HH, and Terwilliger JD (2000a) Linkage analysis in the presence of errors III: marker loci and their map as nuisance parameters. *Am J Hum Genet* 66: 1298-1309
- Göring HH, and Terwilliger JD (2000b) Linkage analysis in the presence of errors IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. *Am J Hum Genet* 66: 1310-1327
- Haimila KE, Partanen JA, and Holopainen PM (2002) Genetic polymorphism of the human ICOS gene. *Immunogenetics* 53: 1028-1032
- Hall FC, Bowness P (1996) HLA and disease: from molecular function to disease association? In: Browning M, McMichael A (eds) *HLA and MHC: genes, molecules and function*. BIOS Scientific Publishers Ltd, Oxford, pp 353-381
- Halstensen TS, Scott H, and Brandtzaeg P (1989) Intraepithelial T cells of the TCR g/d+ CD8- and Vd1/Jd1+ phenotypes are increased in coeliac disease. *Scand J Gastroenterol* 30: 665-672
- Halttunen T, and Mäki M (1999) Serum immunoglobulin A from patients with celiac disease inhibits human T84 intestinal crypt epithelial cell differentiation. *Gastroenterology* 116: 566-572
- Hannigan M, Bourke M, Stevens FM, and McCarthy CF (1991) Gm typing of Irish coeliac patients and controls does not help locate the "second" coeliac gene. *Ir J Med Sci* 160: 57-58

## REFERENCES

- Hearne CM, Ghosh S, and Todd JA (1992) Microsatellite for linkage analysis of genetic traits. *Trends in Genetics* 8: 288-294
- Hervonen K, Karell K, Holopainen P, Collin P, Partanen J, and Reunala T (2000) Concordance of dermatitis herpetiformis and celiac disease in monozygous twins. *J Invest Dermatol* 115: 990-993
- Hetzel PA, Bennett GD, Sheldon AB, ProPERT DN, Brown R, Hay JA, Gabb B, Davidson GP, Schanfield MS, and LaBrooy JT (1987) Genetic markers in Australian Caucasian subjects with coeliac disease. *Tissue Antigens* 30: 18-22
- Houlston RS, Tomlinson IP, Ford D, Seal S, Marossy AM, Ferguson A, Holmes GK, Hosie KB, Howdle PD, Jewell DP, Godkin A, Kerr GD, Kumar P, Logan RF, Love AH, Johnston S, Marsh MN, Mitton S, O'Donoghue D, Roberts A, Walker Smith JA, and Stratton MF (1997) Linkage analysis of candidate regions for coeliac disease genes. *Hum Mol Genet* 6: 1335-1339
- Howdle PD, Losowsky MS (1992) Coeliac disease in adults. In: Marsh MN (ed) *Coeliac disease*. Blackwell Scientific Publications, Oxford, pp 49-80
- Huang D, Giscombe R, Zhou Y, Pirskanen R, and Lefvert AK (2000) Dinucleotide repeat expansion in the CTLA-4 gene leads to T cell hyper-reactivity via the CD28 pathway in myasthenia gravis. *J Neuroimmunol* 105: 69-77
- Ihara K, Ahmed S, Nakao F, Kinukawa N, Kuromaru R, Matsuura N, Iwata I, Nagafuchi S, Kohno H, Miyako K, and Hara T (2001) Association studies of CTLA-4, CD28, and ICOS gene polymorphisms with type 1 diabetes in the Japanese population. *Immunogenetics* 53: 447-454
- Iltanen S, Collin P, Korpela M, Holm K, Partanen J, Polvi A, and Mäki M (1999) Celiac disease and markers of celiac disease latency in patients with primary Sjögren's syndrome. *Am J Gastroenterol* 94: 1042-1046
- Ivarsson A, Persson LA, Nyström L, Ascher H, Cavell B, Danielsson L, Dannaeus A, Lindberg T, Lindquist B, Stenhammar L, and Hernell O (2000) Epidemic of coeliac disease in Swedish children. *Acta Paediatr* 89: 165-171
- Ivarsson SA, Carlsson A, Bredberg A, Alm J, Aronsson S, Gustafsson J, Hagenas L, Hager A, Kristrom B, Marcus C, Moell C, Nilsson KO, Tuvemo T, Westphal O, Albertsson-Wikland K, and Aman J (1999) Prevalence of coeliac disease in Turner syndrome. *Acta Paediatr* 88: 933-936
- Jeffreys AJ, Wilson V, and Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature* 314: 67-73
- Jin L, Macaubas C, Hallmayer J, Kimura A, and Mignot E (1996) Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. *Proc Natl Acad Sci USA* 93: 15285-15288
- Johnston SD, Watson RG, McMillan SA, Sloan J, and Love AH (1997) Prevalence of coeliac disease in Northern Ireland. *Lancet* 350: 1370
- Kan YW, and Dozy AM (1978) Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proc Natl Acad Sci USA* 75: 5631-5635

## REFERENCES

- Karell K, Holopainen P, Mustalahti K, Collin P, Mäki M, and Partanen J (2002) Not all HLA DR3 DQ2 haplotypes confer equal susceptibility to coeliac disease: transmission analysis in families. *Scand J Gastroenterol* 37: 56-61
- Kauppi P, Lindblad-Toh K, Sevón P, Toivonen HT, Rioux JD, Villapakkam A, Laitinen LA, Hudson TJ, Kere J, and Laitinen T (2001) A second-generation association study of the 5q31 cytokine gene cluster and the interleukin-4 receptor in asthma. *Genomics* 77: 35-42
- Kere J (2001) Human population genetics: lessons from Finland. *Annu Rev Genomics Hum Genet* 2: 103-128
- Keuning JJ, Pena AS, van Leeuwen A, van Hooff JP, and van Rood JJ (1976) HLA-Dw3 associated with coeliac disease. *Lancet* 1: 506-508
- King AL, Fraser JS, Moodie SJ, Curtis D, Dearlove AM, Ellis HJ, Rosen-Bronson S, and Ciclitira PJ (2001) Coeliac disease: follow-up linkage study provides further support for existence of a susceptibility locus on chromosome 11p11. *Ann Hum Genet* 65: 377-386
- King AL, Moodie SJ, Fraser JS, Curtis D, Reid E, Dearlove AM, Ellis HJ, and Ciclitira PJ (2002) CTLA-4/CD28 gene region is associated with genetic susceptibility to coeliac disease in UK families. *J Med Genet* 39: 51-54
- King AL, Yiannakou JY, Brett PM, Curtis D, Morris MA, Dearlove AM, Rhodes M, Rosen-Bronson S, Mathew C, Ellis HJ, and Ciclitira PJ (2000) A genome-wide family-based linkage study of coeliac disease. *Ann Hum Genet* 64: 479-490
- Kolho KL, Färkkilä M, and Savilahti E (1998) Undiagnosed coeliac disease is common in Finnish adults. *Scand J Gastroenterol* 33: 1280-1283
- Korponay-Szabo IR, Kovacs JB, Czinner A, Goracz G, Vamos A, and Szabo T (1999) High prevalence of silent celiac disease in preschool children screened with IgA/IgG antiendomysium antibodies. *J Pediatr Gastroenterol Nutr* 28: 26-30
- Kouki T, Sawai Y, Gardine C, Fisfalen M-E, Alegre M-L, and DeGroot LJ (2000) CTLA-4 gene polymorphism at position 49 in exon 1 reduces the inhibitory function of CTLA-4 and contributes to the pathogenesis of Graves' disease. *J Immunol* 165: 6606-6611
- Kristiansen OP, Larsen ZM, and Pociot F (2000) CTLA-4 in autoimmune diseases - a general susceptibility gene to autoimmunity? *Genes and Immunity* 1: 170-184
- Kruglyak L (1997) The use of a genetic map of biallelic markers in linkage studies. *Nat Genet* 17: 21-24
- Kruglyak L, Daly MJ, Reeve-Daly MP, and Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58: 1347-1363
- Kruglyak L, and Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 57: 439-454
- Laitinen T, Daly MJ, Rioux JD, Kauppi P, Laprise C, Petays T, Green T, Cargill M, Haahtela T, Lander ES, Laitinen LA, Hudson TJ, and Kere J (2001) A susceptibility locus for asthma-related traits on chromosome 7 revealed by genome-wide scan in a founder population. *Nat Genet* 28: 87-91

## REFERENCES

- Lander ES, and Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11: 241-247
- Lander ES, and Schork NJ (1994) Genetic dissection of complex traits. *Science* 265: 2037-2048
- Lee KM, Chuang E, Griffin M, Khattri R, Hong DK, Zhang W, Straus D, Samelson LE, Thompson CB, and Bluestone JA (1998) Molecular basis of T cell inactivation by CTLA-4. *Science* 282: 2263-2266
- Lenschow DJ, Walunas TL, and Bluestone JA (1996) CD28/B7 system of T cell costimulation. *Annu Rev Immunol* 14: 233-258
- Lepore L, Martellosi S, Pennesi M, Falcini F, Ermini ML, Ferrari R, Perticarari S, Presani G, Lucchesi A, Lapini M, and Ventura A (1996) Prevalence of celiac disease in patients with juvenile chronic arthritis. *J Pediatr* 129: 311-313
- Lie BA, Sollid LM, Ascher H, Ek J, Akselsen HE, Ronningen KS, Thorsby E, and Undlien DE (1999) A gene telomeric of the HLA class I region is involved in predisposition to both type 1 diabetes and coeliac disease. *Tissue Antigens* 54: 162-168
- Lieberman AP, and Fischbeck KH (2000) Triplet repeat expansion in neuromuscular disease. *Muscle Nerve* 23: 843-850
- Ligers A, Teleshova N, Masterman T, Huang WX, and Hillert J (2001) CTLA-4 gene expression is influenced by promoter and exon 1 polymorphisms. *Genes Immun* 2: 145-152
- Lind S, Eriksson M, Rystedt E, Wiklund O, Angelin B, and Eggertsen G (1998) Low frequency of the common Norwegian and Finnish LDL-receptor mutations in Swedish patients with familial hypercholesterolaemia. *J Intern Med* 244: 19-25
- Ling V, Wu PW, Finnerty HF, Agostino MJ, Graham JR, Chen S, Jussiff JM, Fisk GJ, Miller CP, and Collins M (2001) Assembly and annotation of human chromosome 2q33 sequence containing the CD28, CTLA4, and ICOS gene cluster: analysis by computational, comparative, and microarray approaches. *Genomics* 78: 155-168
- Linsley PS (2001) T cell activation: you can't get good help. *Nat Immunol* 2: 139-140
- Litt M, and Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 44: 397-401
- Logan RF (1992) Epidemiology of coeliac disease. In: Marsh MN (ed) *Coeliac disease*. Blackwell Scientific Publications, Oxford, pp 192-214
- Lonjou C, Barnes K, Chen H, Cookson WO, Deichmann KA, Hall IP, Holloway JW, Laitinen T, Palmer LJ, Wjst M, and Morton NE (2000) A first trial of retrospective collaboration for positional cloning in complex inheritance: assay of the cytokine region on chromosome 5 by the consortium on asthma genetics (COAG). *Proc Natl Acad Sci USA* 97: 10942-10947
- Lundin KEA, Scott H, Fausa O, Thorsby E, and Sollid L (1994) T cells from the small intestinal mucosa of a DR4, DQ7/ DR4,DQ8 celiac disease patient preferentially recognize gliadin when presented by DQ8. *Hum Immunol* 41: 285-291

## REFERENCES

- Lundin KEA, Scott H, Hansen T, Paulsen G, Halstensen TS, Fausa O, Thorsby E, and Sollid LM (1993) Gliadin-specific, HLA-DQ( $\alpha 1^*0501$ ,  $\beta 1^*0201$ ) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. *J Exp Med* 178: 187-196
- Luross JA, and Williams NA (2001) The genetic and immunopathological processes underlying collagen-induced arthritis. *Immunology* 103: 407-416
- Maiuri L, Ciacci C, Raia V, Vacca L, Ricciardelli I, Raimondi F, Auricchio S, Quarantino S, and Londei M (2001) FAS engagement drives apoptosis of enterocytes of coeliac patients. *Gut* 48: 418-424
- Markianos K, Daly MJ, and Kruglyak L (2001) Efficient multipoint linkage analysis through reduction of inheritance space. *Am J Hum Genet* 68: 963-977
- Marks J, Shuster S, and Watson AJ (1966) Small-bowel changes in dermatitis herpetiformis. *Lancet* 2: 1280-1282
- Marsh MN (1992) Gluten, major histocompatibility complex, and the small intestine. *Gastroenterology* 102: 330-354
- Mazzola G, Berrino M, Bersanti M, D'Alfonso S, Cappello N, Bottaro A, Curtioni ES, Fusco P, Vallati M, and Bundino S (1992) Immunoglobulin and HLA-DP genes contribute to the susceptibility to juvenile dermatitis herpetiformis. *Eur J Immunogenet* 19: 129-139
- McIntyre LM, Martin ER, Simonsen KL, and Kaplan NL (2000) Circumventing multiple testing: a multilocus Monte Carlo approach to testing for association. *Genet Epidemiol* 19: 18-29
- Mearin ML, Biemond I, Peña AS, Polanco I, Vazquez C, Schreuder GTHM, de Vries RRP, and van Rood JJ (1983) HLA-DR phenotypes in Spanish coeliac children: their contribution to the understanding of the genetics of the disease. *Gut* 24: 532-537
- Meddeb-Garnaoui A, Zeliszewski D, Mougnot JF, Djilali-Saiah I, Caillat-Zucman S, Dormoy A, Gaudebout C, Tongio MM, Baudon JJ, and Sterkers G (1995) Reevaluation of the relative risk for susceptibility to celiac disease of HLA-DRB1, -DQA1, -DQB1, -DPB1, and -TAP2 alleles in a French population. *Hum Immunol* 43: 190-199
- Meeuwisse GW (1970) Diagnostic criteria in coeliac disease. *Acta Paediatr Scand* 59: 461-463
- Meloni G, Dore A, Fanciulli G, Tanda F, and Bottazzo GF (1999) Subclinical coeliac disease in school-children from northern Sardinia. *Lancet* 353: 37
- Molberg O, Kett K, Scott H, Thorsby E, Sollid LM, and Lundin KE (1997) Gliadin specific, HLA DQ2-restricted T cells are commonly found in small intestinal biopsies from coeliac disease patients, but not from controls. *Scand J Immunol* 46: 103-109
- Molberg O, McAdam SN, Körner R, Quarsten H, Kristiansen C, Madsen L, Fugger L, Scott H, Noren O, Roepstorff P, Lundin KEA, Sjöström H, and Sollid LM (1998) Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat Med* 4: 713-717
- Molberg O, McAdam SN, and Sollid LM (2000) Role of tissue transglutaminase in celiac disease. *J Pediatr Gastroenterol Nutr* 30: 232-240
- Morel L, Tian XH, Croker BP, and Wakeland EK (1999) Epistatic modifiers of autoimmunity in a murine model of lupus nephritis. *Immunity* 11: 131-139

## REFERENCES

- Morgan TH (1911) Random segregation versus coupling in mendelian inheritance. *Science* 34: 384
- Morris MA, Yiannakou JY, King AL, Brett PM, Biagi F, Vaughan R, Curtis D, and Ciclitira PJ (2000) Coeliac disease and Down syndrome: associations not due to genetic linkage on chromosome 21. *Scandinavian Journal of Gastroenterology* 35: 177-180
- Morton NE (1955) Sequential tests for the detection of linkage. *Am J Hum Genet* 7: 277-318
- Morton NE (1956) The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. *Am J Hum Genet* 8: 80-96
- Mustalahti K, Sulkanen S, Holopainen P, Laurila K, Collin P, Partanen J, and Mäki M (2002) Coeliac disease among healthy members of multiple case coeliac disease families. *Scand J Gastroenterol* 37: 161-165
- Mäki M (1995) The humoral immune system in coeliac disease. *Baillieres Clin Gastroenterol* 9: 231-249
- Mäki M, and Collin P (1997) Coeliac disease. *Lancet* 349: 1755-1759
- Mäki M, Holm K, Lipsanen V, Hällström O, Viander M, Collin P, Savilahti E, and Koskimies S (1991) Serological markers and HLA genes among healthy first-degree relatives of patients with coeliac disease. *Lancet* 338: 1350-1353
- Mäki M, Kallonen K, Lähdeaho ML, and Visakorpi JK (1988) Changing pattern of childhood coeliac disease in Finland. *Acta Paediatr Scand* 77: 408-412
- Naluai TA, Nilsson S, Gudjonsdottir AH, Louka AS, Ascher H, Ek J, Hallberg B, Samuelsson L, Kristiansson B, Martinsson T, Nerman O, Sollid LM, and Wahlstrom J (2001) Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for celiac disease on chromosomes 5 and 11. *Eur J Hum Genet* 9: 938-944
- Neuhausen SL, Feolo M, Farnham J, Book L, and Zone JJ (2001) Linkage analysis of HLA and candidate genes for celiac disease in a North American family-based study. *BMC Med Genet* 2: 12
- Nisticó L, Buzzetti R, Pritchard LE, and et al (1996) The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. *Hum Mol Genet* 5: 1075-1080
- Not T, Horvath K, Hill ID, Partanen J, Hammed A, Magazzu G, and Fasano A (1998) Celiac disease risk in the USA: High prevalence of antiendomysium Antibodies in healthy blood donors. *Scand J Gastroenterol* 33: 494-498
- Nowotny P, Kwon JM, and Goate AM (2001) SNP analysis to dissect human traits. *Curr Opin Neurobiol* 11: 637-641
- O'Connell JR, and Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63: 259-266
- Ose L (1999) An update on familial hypercholesterolaemia. *Ann Med* 31 Suppl 1:13-8.: 13-18
- Palmer L, Lonjou C, Barnes K, Chen H, Cookson WO, Deichmann KA, Holloway JW, Laitinen T, Wjst M, and Morton NE (2001) A retrospective collaboration on chromosome 5 by the International Consortium on Asthma Genetics (COAG). *Clin Exp Allergy* 31: 152-154

## REFERENCES

- Partanen J (2000) The HLA-DRB4 gene does not explain genetic susceptibility in HLA-DQ2-negative celiac disease. *Immunogenetics* 51: 249-251
- Partanen J, Westman P (1997) Caucasian Finnish normal. In: Terasaki PI, Gjertson DW (eds) *HLA 1997*. UCLA Tissue Typing Laboratory, Los Angeles, pp 214-215
- Peltonen L, Jalanko A, and Varilo T (1999) Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* 10: 1913-1923
- Pender SL, Tickle SP, Docherty AJ, Howie D, Wathen NC, and MacDonald TT (1997) A major role for matrix metalloproteinases in T cell injury in the gut. *J Immunol* 158: 1582-1590
- Perola M, Kainulainen K, Pajukanta P, Terwilliger JD, Hiekkalinna T, Ellonen P, Kaprio J, Koskenvuo M, Kontula K, and Peltonen L (2000) Genome-wide scan of predisposing loci for increased diastolic blood pressure in Finnish sibs. *J Hypertens* 18: 1579-1585
- Petronzelli F, Bonamico M, Ferrante P, Grillo R, Mora B, Mariani P, Apollonio I, Gemme G, and Mazzilli MC (1997) Genetic contribution of the HLA region to the familial clustering of coeliac disease. *Ann Hum Genet* 61: 307-317
- Ploski R, Ek J, Thorsby E, and Sollid LM (1993) On the HLA-DQ( $\alpha$ 1\*0501,  $\beta$ 1\*0201)-associated susceptibility in celiac disease: a possible gene dosage effect of DQB1\*0201. *Tissue Antigens* 41: 173-177
- Polanco I, Biemond I, van Leeuwen A, Schreuder I, Kahn PM, Guerrero J, D'Amaro J, Vasquez C, van Rood JJ, Peña AS (1981) Gluten sensitive enteropathy in Spain: genetic and environmental factors. In: McConnell RB (ed) *The genetics of coeliac disease*. MTP, Lancaster, pp 211-231
- Polvi A, Arranz E, Fernández-Arquero M, Collin P, Mäki M, Sanz A, Calvo C, Maluenda C, Westman P, de la Concha EG, and Partanen J (1998) HLA-DQ2-negative celiac disease in Finland and Spain. *Hum Immunol* 59: 169-175
- Polvi A, Eland C, Koskimies S, Mäki M, and Partanen J (1996) HLA DQ and DP in Finnish families with celiac disease. *Eur J Immunogenet* 23: 221-234
- Polvi A, Garden OA, Houlston RS, Mäki M, Batt RM, and Partanen J (1998) Genetic susceptibility to gluten sensitive enteropathy in Irish setter dogs is not linked to the major histocompatibility complex. *Tissue Antigens* 52: 543-549
- Polvi A, Mäki M, and Partanen J (1997) Celiac patients predominantly inherit HLA-DPB1\*0101 positive haplotype from HLA-DQ2 homozygous parent. *Hum Immunol* 53: 156-158
- Polymeropoulos MH, Xiao H, Rath DS, and Merrill CR (1991) Dinucleotide repeat polymorphism at the human CTLA4 gene. *Nucleic Acids Res* 19: 4018
- Popat S, Hearle N, Wixey J, Hogberg L, Bevan S, Lim W, Stenhammar L, and Houlston RS (2002) Analysis of the CTLA4 gene in Swedish coeliac disease patients. *Scand J Gastroenterol* 37: 28-31
- Rautonen N, Rautonen J, and Savilahti E (1990) Influence of the G2m(n) allotype and age on IgG subclass distribution in antibodies to dietary proteins in children with coeliac disease. *Clin Exp Immunol* 81: 306-310
- Reunala T (1996) Incidence of familial dermatitis herpetiformis. *Br J Dermatol* 134: 394-398

## REFERENCES

- Reunala T, Kosnai I, Karpatai S, Kuitunen P, Török E, and Savilahti E (1984) Dermatitis herpetiformis: jejunal findings and skin response to gluten-free diet. *Arch Dis Child* 59: 517-522
- Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, Delmonte T, Kocher K, Miller K, Guschwan S, Kulbokas EJ, 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 SB, McLeod RS, Griffiths AM, Bitton A, Greenberg GR, Lander ES, Siminovitch KA, and Hudson TJ (2001) Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 29: 223-228
- Risch N (1987) Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet* 40: 1-14
- Risch NJ (2000) Searching for genetic determinants in the new millennium. *Nature* 405: 847-856
- Roschmann E, Wienker TF, Gerok W, and Volk BA (1993) T-cell receptor variable genes and genetic susceptibility to celiac disease: an association and linkage study. *Gastroenterology* 105: 1790-1796
- Roschmann E, Wienker TF, Gerok W, and Volk BA (1995) Analysis of marker genes contributing to coeliac disease susceptibility. *Adv Exp Med Biol* 371B: 1339-1343
- Roschmann E, Wienker TF, and Volk BA (1996) Role of T cell receptor delta gene in susceptibility to celiac disease. *J Mol Med* 74: 93-98
- Rotter JI, and Landaw EM (1984) Measuring the genetic contribution of a single locus to a multilocus disease. *Clinical Genetics* 26: 529-542
- Sategna-Guidetti C, Bruno M, Mazza E, Carlino A, Predebon S, Tagliabue M, and Brossa C (1998) Autoimmune thyroid diseases and coeliac disease. *Eur J Gastroenterol Hepatol* 10: 927-931
- Sategna-Guidetti C, Solerio E, Scaglione N, Aimo G, and Mengozzi G (2001) Duration of gluten exposure in adult coeliac disease does not correlate with the risk for autoimmune disorders. *Gut* 49: 502-505
- Savilahti E, Pelkonen P, Verkasalo M, and Koskimies S (1985) Selective deficiency of immunoglobulin A in children. *Klin Padiatr* 197: 336-340
- Savilahti E, Pelkonen P, and Visakorpi JK (1971) IgA deficiency in children. A clinical study with special reference to intestinal findings. *Arch Dis Child* 46: 665-670
- Savilahti E, Reunala T, and Mäki M (1992) Increase of lymphocytes bearing the gamma/delta T cell receptor in the jejunum of patients with dermatitis herpetiformis. *Gut* 33: 206-211
- Schaid DJ, and Sommer SS (1994) Comparison of statistics for candidate-gene association studies using cases and parents. *Am J Hum Genet* 55: 402-409
- Schmitz J (1992) Coeliac disease in childhood. In: Marsh MN (ed) *Coeliac disease*. Blackwell Scientific Publications, Oxford, pp 17-48
- Schneider, S., Kueffer, J.-M., Roessli, D., and Excoffier, L. Arlequin ver 1.1: A software for population genetic data analysis. (1.1). 1997. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

## REFERENCES

- Schuppan D (2000) Current concepts of celiac disease pathogenesis. *Gastroenterology* 119: 234-242
- Scott L, and Rogus JJ (2000) Using unaffected child trios to test for transmission distortion. *Genet Epidemiol* 19: 381-394
- She JX, and Marron MP (1998) Genetic susceptibility factors in type 1 diabetes: linkage, disequilibrium and functional analyses. *Curr Opin Immunol* 10: 682-689
- Shih MC, and Whittemore AS (2001) Allele-sharing among affected relatives: non-parametric methods for identifying genes. *Stat Methods Med Res* 10: 27-55
- Shiner M (1957) Small intestinal biopsies by the oral route. *J Mt Sinai Hosp* 24: 273-277
- Sjöström H, Lundin KE, Molberg O, Körner R, McAdam SN, Anthonsen D, Quarsten H, Noren O, Roepstorff P, Thorsby E, and Sollid LM (1998) Identification of a gliadin T-cell epitope in coeliac disease: general importance of gliadin deamidation for intestinal T-cell recognition. *Scand J Immunol* 48: 111-115
- Smith JB, Tulloch JE, Meyer LJ, and Zone JJ (1992) The incidence and prevalence of dermatitis herpetiformis in Utah. *Arch Dermatol* 128: 1608-1610
- Sollid LM (2000) Molecular basis of celiac disease. *Annu Rev Immunol* 18: 53-81
- Sollid LM, Molberg O, McAdam S, and Lundin KE (1997) Autoantibodies in coeliac disease: tissue transglutaminase—guilt by association? *Gut* 41: 851-852
- Sollid LM, and Thorsby E (1993) HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 105: 910-922
- Spielman RS, and Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 59: 983-989
- Spielman RS, McGinnis RE, and Ewens WJ (1993) Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52: 506-516
- Spurkland A, Sollid LM, Polanco I, Vartdal F, and Thorsby E (1992) HLA-DR and -DQ genotypes of celiac disease patients serologically typed to be non-DR3 or non-DR5/7. *Hum Immunol* 35: 188-192
- Stern M (2000) Comparative evaluation of serologic tests for celiac disease: a European initiative toward standardization. *J Pediatr Gastroenterol Nutr* 31: 513-519
- Stokes PL, Asquith P, Holmes GKT, Mackintosh P, and Cooke WT (1972) Histocompatibility antigens associated with adult coeliac disease. *Lancet* 2: 162-164
- The MHC sequencing consortium (1999) Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401: 921-923
- Thompson CB, and Allison JP (1997) The emerging role of CTLA-4 as an immune attenuator. *Immunity* 7: 445-450

## REFERENCES

- Tivol EA, Borriello F, Schweitzer AN, Lynch WP, Bluestone JA, and Sharpe AH (1995) Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. *Immunity* 3: 541-547
- Torinsson-Nalwai Å, Nilsson S, Samuelsson L, Gudjonsdottir AH, Ascher H, Ek J, Hallberg B, Kristiansson B, Martinsson T, Nerman O, Sollid LM, and Wahlström J (2000) The CTLA4/CD28 gene region on chromosome 2q33 confers susceptibility to celiac disease in a way possibly distinct from that of type 1 diabetes and other chronic inflammatory disorders. *Tissue Antigens* 56: 350-355
- Trabace S, Giunta A, Rosso M, Marzorati D, Cascino I, Tettam A, Mazzilli MC, and Gandini E (1998) HLA-ABC and DR antigens in celiac disease. A study in a pediatric Italian population. *Vox.Sang.*
- Troncone R, and Ferguson A (1991) Anti-gliadin antibodies. *J Pediatr Gastroenterol Nutr* 12: 150-158
- Trowsdale J (1996) Molecular genetics of HLA class I and class II regions. In: Browning M, McMichael A (eds) *HLA and MHC: genes, molecules and function*. BIOS Scientific Publishers Ltd, Oxford, pp 23-38
- van Belzen MJ, Mulder CJ, Pearson PL, Houwen RH, and Wijmenga C (2001) The tissue transglutaminase gene is not a primary factor predisposing to celiac disease. *Am J Gastroenterol* 96: 3337-3340
- van de Wal Y, Kooy YM, van Veelen P, Pena S, Mearin L, Papadopoulos G, and Koning F (1998a) Cutting edge: selective deamination by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. *J Immunol* 158: 1585-1588
- van de Wal Y, Kooy YM, van Veelen PA, Pena SA, Mearin LM, Molberg O, Lundin KE, Sollid LM, Mutis T, Benckhuijsen WE, Drijfhout JW, and Koning F (1998b) Small intestinal T cells of celiac disease patients recognize a natural pepsin fragment of gliadin. *Proc Natl Acad Sci USA* 95: 10050-10054
- Ventura A, Magazzu G, and Greco L (1999) Duration of exposure to gluten and risk for autoimmune disorders in patients with celiac disease. SIGEP Study Group for Autoimmune Disorders in Celiac Disease. *Gastroenterology* 117: 297-303
- Vuorio AF, Aalto-Setälä K, Koivisto UM, Turtola H, Nissen H, Kovanen PT, Miettinen TA, Gylling H, Oksanen H, and Kontula K (2001) Familial hypercholesterolaemia in Finland: common, rare and mild mutations of the LDL receptor and their clinical consequences. Finnish FH-group. *Ann Med* 33: 410-421
- Wandstrat A, and Wakeland E (2001) The genetics of complex autoimmune diseases: non-MHC susceptibility genes. *Nat Immunol* 2: 802-809
- Waterhouse P, Penninger JM, Timms E, Wakeham A, Shahinian A, Lee KP, Thompson CB, Griesser H, and Mak TW (1995) Lymphoproliferative disorders with early lethality in mice deficient in CtlA-4. *Science* 270: 985-988
- Weber JL, and May PE (1989) Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44: 388-396
- Weiss JB, Austin RK, Schanfield MS, and Kagnoff MF (1983) Gluten-sensitive enteropathy. Immunoglobulin G heavy-chain (Gm) allotypes and the immune response to wheat gliadin. *J Clin Invest* 72: 96-101

## REFERENCES

- Whitacre CC, Reingold SC, and O'Looney PA (1999) A gender gap in autoimmunity. *Science* 283: 1277-1278
- White PC, New MI, and Dupont B (1984) HLA-linked congenital adrenal hyperplasia results from a defective gene encoding a cytochrome P-450 specific for steroid 21-hydroxylation. *Proc Natl Acad Sci USA* 81: 7505-7509
- Wicker LS, Todd JA, and Peterson LB (1995) Genetic control of autoimmune diabetes in the NOD mouse. *Annu Rev Immunol* 13: 179-200
- Wise LH, Lanchbury JS, and Lewis CM (1999) Meta-analysis of genome searches. *Ann Hum Genet* 63: 263-272
- Woolley N, Holopainen P, Ollikainen V, Mustalahti K, Mäki M, Kere J, Partanen J (2002) A new locus for coeliac disease mapped to chromosome 15 in a population isolate. *Hum Genet* (in press)
- Wright AF, Carothers AD, and Pirastu M (1999) Population choice in mapping genes for complex diseases. *Nat Genet* 23: 397-404
- Xu J, Postma DS, Howard TD, Koppelman GH, Zheng SL, Stine OC, Bleecker ER, and Meyers DA (2000) Major genes regulating total serum immunoglobulin E levels in families with asthma. *Am J Hum Genet* 67: 1163-1173
- Yiannakou JY, Brett PM, Morris MA, Curtis D, Mathew C, Vaughan R, Rosen-Bronson S, and Ciclitira PJ (1999) Family linkage study of the T-cell receptor genes in coeliac disease. [see comments]. *Italian Journal of Gastroenterology & Hepatology* 31: 198-201
- Yokouchi Y, Nukaga Y, Shibasaki M, Noguchi E, Kimura K, Ito S, Nishihara M, Yamakawa-Kobayashi K, Takeda K, Imoto N, Ichikawa K, Matsui A, Hamaguchi H, and Arinami T (2000) Significant evidence for linkage of mite-sensitive childhood asthma to chromosome 5q31-q33 near the interleukin 12 B locus by a genome-wide search in Japanese families. *Genomics* 66: 152-160
- Zhao H (2000) Family-based association studies. *Stat Methods Med Res* 9: 563-587
- Zhong F, McCombs CC, Olson JM, Elston RC, Stevens FM, McCarthy CF, and Michalski JP (1996) An autosomal screen for genes that predispose to celiac disease in the western counties of Ireland. *Nat Genet* 14: 329-333