Coeliac disease

Investigation of the genetic factors underlying coeliac disease

Coeliakie

Een onderzoek naar de genetische factoren die coeliakie veroorzaken

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de Rector Magnificus, Prof. dr. W.H. Gispen
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen
op dinsdag 16 december 2003 des namiddags te 14.30 uur

door

Martine Juliana van Belzen

Geboren op 15 maart 1973, te Barendrecht

Promotores: Prof. dr. C. Wijmenga[1]
Prof. dr. P.L. Pearson[1]

Copromotor: Dr. R.H.J. Houwen[2]

[1]Department of Biomedical Genetics, University Medical Centre (UMC) Utrecht, the Netherlands.
[2]Department of Paediatric Gastroenterology, UMC Utrecht, the Netherlands.

The picture on the front cover shows two ears of barley in a field of wheat, taken in June 2003 near Groesbeek.

ISBN 90-393-3542-7

The work presented in this thesis was performed at the Department of Biomedical Genetics of the UMC Utrecht, within the Research School of Developmental Biology.

The work presented in this thesis was financially supported by the Netherlands Digestive Diseases Foundation (Maag Lever Darm Stichting (MLDS)) and printing of this thesis was financially supported by Nutricia Nederland B.V., MLDS, the UMC Utrecht and the Department of Biomedical Genetics.

Thesis Committee

Prof. dr. D.E. Grobbee
Julius Centre for Health Sciences and Primary Care, UMC Utrecht, the Netherlands.

Prof. dr. W.T. Kuis
Department of Paediatrics, UMC Utrecht, the Netherlands.

Dr. M. Samson
Department of Gastroenterology, UMC Utrecht, the Netherlands.

Prof. dr. B.A. Oostra
Department of Clinical Genetics, Erasmus Medical Centre, Rotterdam, the Netherlands.

Prof. dr. A.S. Peña
Department of Immunogenetics, Vrije Universiteit Medical Centre, Amsterdam, the Netherlands.

Te weten wat men weet
en te weten wat men niet weet,
dat is kennis
Confucius

voor mijn ouders

Contents

Chapter 1

General Introduction

Coeliac Disease

History

The first description of coeliac disease dates back to 1888, when Gee described a disease with onset usually between one and five years of age with diarrhoea, abdominal distension and failure to thrive as the most important symptoms.[1] In 1908 Herter extended this description with delay in vertical growth.[2] The cause of coeliac disease was unknown at that time, but by the end of the 1920's it was generally agreed that the disease should be treated by rest and diet. Many diets have been recommended, from the banana diet, the carbohydrate diet (fruit, tomatoes and mashed potatoes), beefsteak therapy, and the milk diet (2-2.5 litres/day).[3] But it was not until 1941 that the famous Dutch paediatrician Willem Dicke published his findings about the benefits of a wheat-free diet.[4] At that time, the banana diet and the fruit and vegetables diet were regarded as the best diets for treating coeliac disease. However, during World War II, bananas, fruit and vegetables were not easily available, but Dicke's patients were doing well on a simple diet without bread or rusks. In his thesis, published in 1950, he describes the growth curves and symptoms of five children who clinically improved when wheat, rye and oats flour were omitted from their diet, and who relapsed when these flours were added to the diet again, thereby proving that components from these grains cause coeliac disease.[5]

It is now well established that one of the major protein components of wheat, the storage proteins named gluten, are the toxic components that cause coeliac disease. Gluten proteins contain high percentages of glutamine and proline residues, hence the name "gluten". Wheat gluten can be divided into two groups, the gliadins which are alcohol-soluble and the glutenins which are insoluble. Similar storage proteins are present in barley, rye and oats, and are called hordeins, secalins and avenins, respectively. Wheat is the most toxic for coeliac disease patients. Barley and rye, closely related in evolutionary terms, are also harmful but more distantly related grains like corn and rice are non-toxic.[6] Oats are evolutionary categorised between these groups and have long been regarded as harmful. Recent studies however, have shown that oats can be tolerated well by coeliac disease patients, when they are not contaminated by gluten proteins from wheat, barley or rye.[7] Coeliac disease is still treated with a life-long gluten-free diet, a diet without wheat, barley or rye products.

Clinical aspects

Ingestion of gluten by coeliac disease patients results in lesions of the proximal small intestine. The range of abnormalities in the small intestinal mucosa can be classified according the modified Marsh classification.[6] Marsh I comprises normal mucosal architecture with a marked infiltration of the villous epithelium by lymphocytes. Marsh II includes intraepithelial lymphocytosis and crypt hyperplasia with branching and elongation of crypts. And Marsh III comprises villous atrophy with lymphocytic infiltration of the epithelium and crypt hyperplasia. This category has been modified into three subtypes, with MIIIa representing partial, MIIIb subtotal and MIIIc total villous atrophy, respectively.[8] Histological identification of the small-intestinal lesion by a biopsy of the duodenum is still the only accepted basis for diagnosing coeliac disease.[9] However, in recent years, serological tests have become available to screen for this disorder. The best screening test available at the moment is the determination of anti-endomysium (Ema) IgA antibodies, with a high sensitivity (85-98%) and specificity (97-100%) in detecting untreated coeliac disease patients.[9] A few years ago, the enzyme tissue transglutaminase (tTG) was identified as being the autoantigen recognized in the endomysium.[10] The sensitivity of the anti-tTG IgA antibody test seems to be as high as that of the IgA-anti-endomysium antibody test (90-98%), but its specificity is slightly lower (94-97%).[9] However, the sensitivity of both antibody tests was shown to be lower in patients with only partial villous atrophy.[8,11]

The clinical presentation of coeliac disease comprises a wide spectrum of symptoms, most of them related to malabsorption of nutrients from food.[9] Typical symptoms of childhood coeliac disease include chronic diarrhoea, abdominal distension and a failure to thrive. However, the occurrence of the classic presentation of coeliac disease in infancy has decreased. Nowadays, children with coeliac disease may present at a later age with isolated short stature, recurrent abdominal pain or with delayed puberty. Recurrent aphthous lesions in the mouth, dental enamel defects, fatigue, isolated iron deficiency anaemia, or dermatitis herpetiformis, characterized by itching skin lesions, may also be manifestations of coeliac disease, both in children and adults. Moreover, infertility or recurrent abortions have been observed in women with coeliac disease. Severe complications, like osteoporosis, autoimmune disorders or intestinal malignancies, may occur in patients with untreated coeliac disease.[12-15] Besides the paucisymptomatic

presentation of coeliac disease, some patients are even asymptomatic. Strict adherence to the gluten-free diet results in complete restoration of the small intestine and disappearance of the clinical symptoms.

In addition, coeliac disease is associated with other autoimmune disorders. Coeliac disease is diagnosed in ~4.5% of patients with type 1 diabetes mellitus.[16] Moreover, an apparent association exists between thyroid disease and coeliac disease, as approximately 5% of patients with autoimmune thyroiditis are found to be positive for coeliac disease after screening. Similarly, about 5% of patients with coeliac disease suffer from autoimmune thyroid disease.[17] Coeliac disease also has an increased prevalence in patients with autoimmune liver diseases[18] and certain chromosomal abnormalities like Down syndrome (3-17%)[19-22] and in women with Turner syndrome (4-6%).[23,24]

Epidemiology

Coeliac disease occurs largely in Caucasians. Although the disease has been well documented in Asians from India[25] and Turkey[26], it is rare or nonexistent among native Africans, Japanese and Chinese. The disease is less common in men than in women, with a male to female ratio of 1:3. Up to only a few years ago, coeliac disease was thought to be rare; most patients were not recognized since they did not have the symptoms associated with classical coeliac disease. In the Netherlands, the incidence of recognized childhood coeliac disease has increased from 0.18/1000 live births in the period 1975-1990, to 0.54/1000 live births in 1994.[27] In 1998, two large population screenings were conducted in the Netherlands. Over 6000 children[28] and 1000 blood donors[29] were screened for the presence of Ema-antibodies. Coeliac disease was confirmed by small-bowel biopsy in Ema-positive individuals. These studies revealed that the prevalence of coeliac disease is 1 in 200-300, which is much higher than was previously thought. Furthermore, it was estimated that for every child recognized with coeliac disease, 14 affected children go unrecognized.[28] The disease prevalence in the Netherlands is comparable with that seen in other European countries[30-32], Australia[33], South America[34,35] and the USA.[36] Coeliac disease is, with an average disease prevalence of approximately 0.4%, one of the most common forms of food intolerance in the world.

Genetics

The development of coeliac disease is influenced by both genetic and environmental factors. It has long been known that coeliac disease runs in families. The sibling recurrence risk is approximately 10%.[6,37-44] So, based on a population prevalence of 0.4%, the sibling relative risk (λ_s) for development of coeliac disease is 25. Evidence for a genetic contribution to coeliac disease was also shown by a recent twin study.[45] Disease concordance rates for coeliac disease were 86% in monozygous (MZ) twins and 20% in dizygous (DZ) twins. Although the large difference between MZ twins and DZ twins indicates a strong genetic component in the development of coeliac disease, environmental factors are also of vital importance as can be concluded from the less than 100% concordance rate in MZ twins.

To date only one genetic factor involved in coeliac disease is known, namely the HLA-DQ protein. The HLA-DQ molecule consists of two different peptides, HLA-DQα1 and HLA-DQβ1. These peptides are encoded by the highly polymorphic HLA-DQA1 and HLA-DQB1 genes respectively, localized in the major histocompatibility complex (MHC) on chromosome region 6q21.3 (Figure 1a). The high degree of polymorphism in both genes and the combination of the two peptides in the heterodimer HLA-DQ result in a large number of possible different DQ molecules. This mechanism is used by the immune system to maximize the number of peptides that can be presented to the T cells. The HLA-DQ2 molecule, encoded by the HLA-DQA1*05 and HLA-DQB1*02 alleles in either the *cis* or the *trans* configuration (Figure 1b), is expressed by more than 90% of coeliac disease patients. This is in strong contrast to the frequency of HLA-DQ2 carriers in the general population, which is 20-30%. Almost all coeliac disease patients carry the alleles encoding HLA-DQ2 on the extended HLA-B8-DR3-DQ2 haplotype. This haplotype includes many other genes that play a role in the immune response and it cannot be excluded that another MHC gene also confers increased risk to coeliac disease development. Almost all the DQ2-negative coeliac disease patients carry the HLA-DQA1*0301 and HLA-DQB1*0302 alleles, which combine into the heterodimer HLA-DQ8.[46]

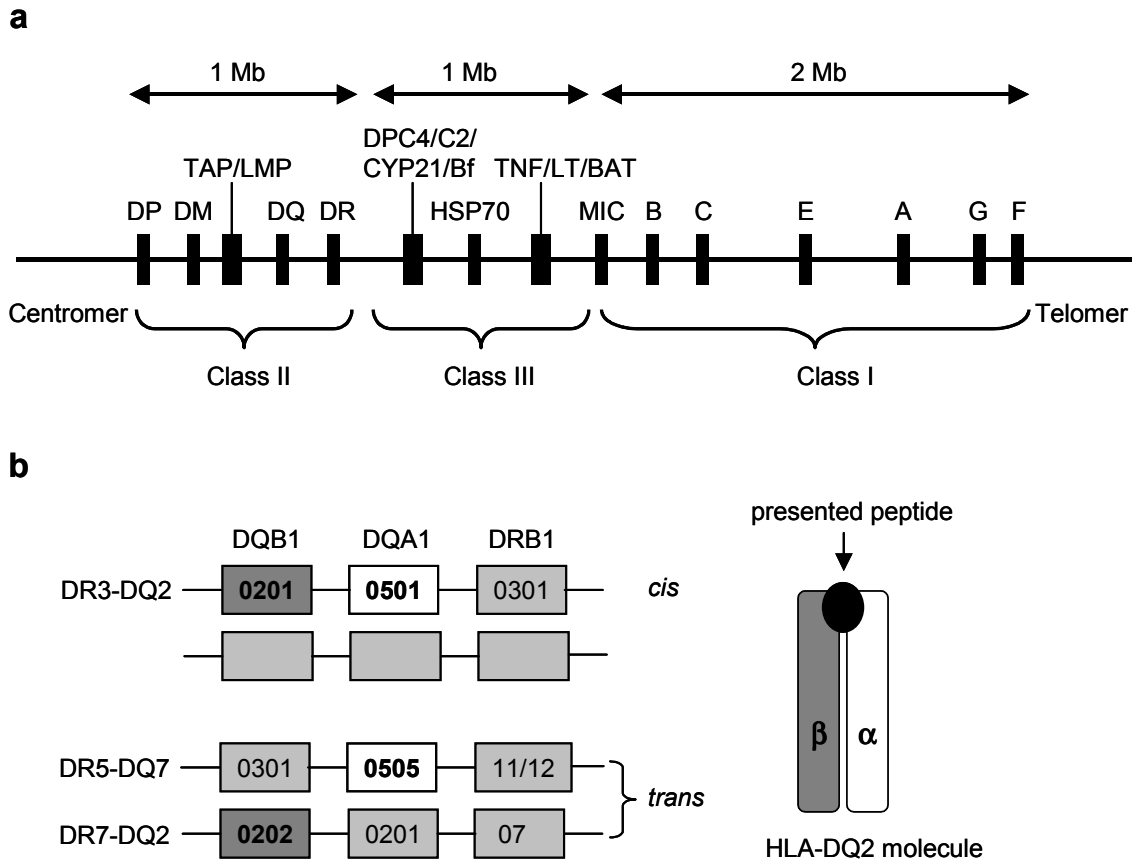


Figure 1. The MHC gene complex and HLA association in coeliac disease.

a. Overview of the most important loci in the MHC complex. The MHC complex contains in total 224 genes, of which 96 are pseudogenes. **b.** The HLA-DQ heterodimer is encoded by the DQA1 and DQB1 genes. Coeliac disease is associated with HLA-DQ2, which is encoded by the HLA-DQA1*05 and HLA-DQB1*02 alleles. These two alleles can be present either in *cis* on the DR3 haplotype or in *trans* on the DR5 and DR7 haplotypes. Both situations result in expression of the HLA-DQ2 molecule.

Coeliac disease is a typical example of a multifactorial disorder, i.e. a disease caused by the combined action of several genes and environmental factors. The genetic predisposition cannot be derived from one single gene. There are several reasons to assume this. First, there is no Mendelian inheritance of the disease in families. Consequently, the recurrence risk of 10% is much lower than would be expected for a recessive or dominant disease (25% or 50%, respectively). Second, the frequency of HLA-DQ2 carriers in the population (20-30%) is much higher than the prevalence of coeliac disease (0.4%), implying that most HLA-DQ2 positive individuals do not develop coeliac disease. And third, the disease concordance rate between HLA identical siblings is approximately 30%,[47] which is much lower than the 86% concordance rate in MZ twins.

Since the HLA identical siblings share at least part of their environment, this difference must be explained by the presence of other susceptibility genes.

Environmental factors

In addition to the genetic contribution, several non-genetic factors play a role in the aetiology of coeliac disease too. The most important environmental factor is gluten. When a coeliac disease patient does not ingest gluten, the disease process is absent and the gut of the patient is completely normal. The effect of gluten intake on coeliac disease development became clear from a recent study on incidence rates in Sweden.[48] The annual incidence rate of coeliac disease was 50-60 cases per 100,000 person years in children under the age of two from 1973 to 1984. In the years from 1985 to 1987, this incidence increased dramatically to 200-240 cases per 100,000 person years. After 1995, the incidence dropped to the same level as before 1985. This sudden increase in coeliac disease incidence was thought to be caused in part by a higher intake of gluten during the years with high incidence rates, compared to the years with low incidence rates. In addition, at the end of 1982 the national recommendation for introduction of gluten at the age of 4 months was postponed to 6 months of age. In 1996, this recommendation was changed to introducing small amounts of gluten at 4 months, preferably while the child is still being breastfed. A new study was performed during the peak years of the Swedish coeliac disease epidemic.[49] The risk of coeliac disease in children under two years of age was reduced when small amounts of gluten were introduced during the breastfeeding period. This risk was even smaller when breast-feeding was continued after introduction of gluten. A higher risk was observed when large amounts of gluten were introduced. These studies suggest that the amount of gluten consumption, breastfeeding habits and timing of gluten introduction can all influence the development of the disease.

Although gluten is essential for the expression of coeliac disease, there must be other environmental factors that influence the development of the disease. The 14% discordance rate in MZ twins cannot be explained by the presence of gluten, as twins are likely to consume comparable diets. The identity of the other environmental factors is not known yet, but infections or the occurrence of other autoimmune diseases are thought to be able to trigger coeliac disease development. Furthermore, children born in the summer seem to have an increased risk for coeliac disease compared to children born

in other seasons.[50] But before the environmental contribution to coeliac disease can be completely unravelled, the genetic contribution and the pathogenesis of the disease must be known.

Pathogenesis

In the last few years, great progress has been made in understanding the pathogenesis of coeliac disease. It has become clear that ingestion of gluten by coeliac disease patients results in an improper T cell mediated immune response. Gluten peptides are not completely digested by the enzymes of the stomach and small-intestinal brush border.[51] The partially digested gluten peptides cross the gut epithelium by an unknown mechanism and arrive in the lamina propria (Figure 2). These peptides are bound by HLA-DQ2 and DQ8 proteins on the cell surface of antigen presenting cells (APCs) and presented to the T cells in the lamina propia.[52] Other HLA-DQ molecules have different binding motifs and are probably incapable of binding these peptides. Interaction between gluten peptides and HLA-DQ2 and DQ8 is greatly enhanced by the action of the enzyme tissue transglutaminase (tTG).[53,54] This enzyme is expressed at the epithelial brush border and is also excreted in the subepithelial region in active coeliac disease.[54] An important function of tTG is to catalyse protein cross linking by the formation of isopeptide bonds between lysine and glutamine residues, resulting in a dense protein network necessary for repair of tissue damage. Alternatively, the tTG enzyme specifically deamidates certain glutamine residues from gluten peptides into glutamic acid. This results in peptides which match perfectly to the binding motif of HLA-DQ2 and DQ8, because these molecules prefer negative charges in the bound peptides (i.e. glutamic acid).[55] Recently, the specificity of tTG was elucidated.[56] It was shown that the amino acid composition following the glutamine residues determines which of these residues will be converted to glutamic acid. From this pattern it could be predicted that glutamine residues from gluten-related peptides derived from oats would not be a target for tTG. This would explain the non-toxicity of oats for coeliac disease patients.

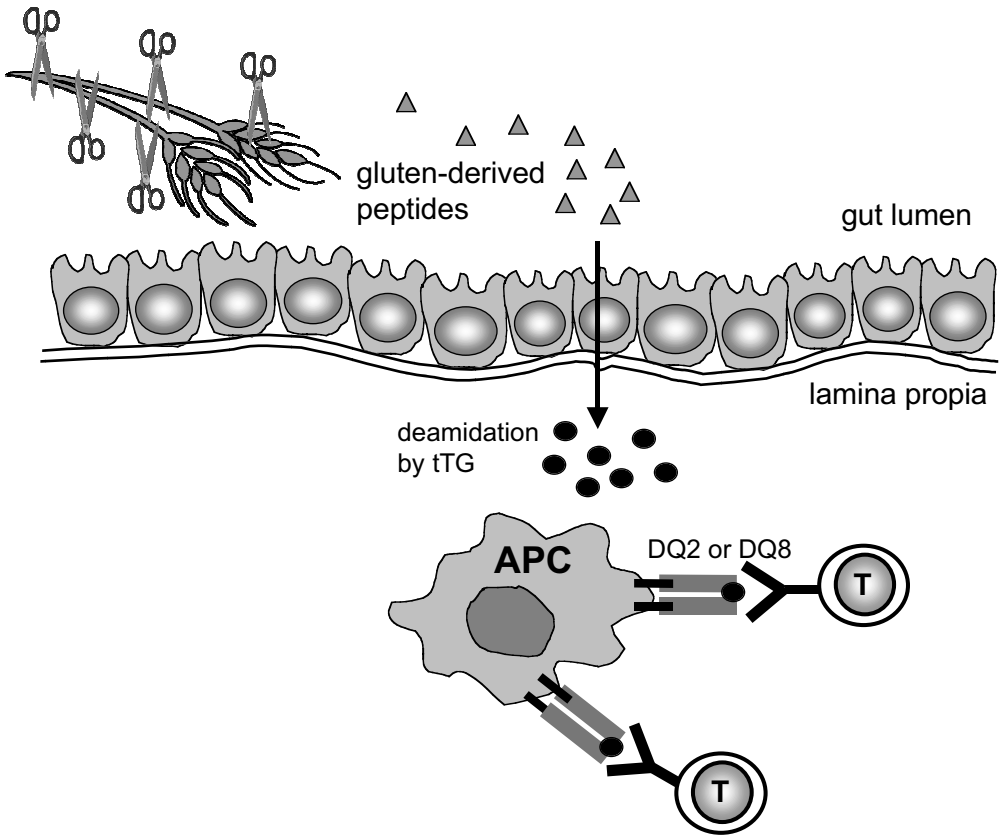


Figure 2. Pathogenesis of coeliac disease.

Gluten peptides are partially digested by the enzymes of the stomach and small-intestinal brush border. The gluten-derived peptides cross the epithelium of the gut and are deamidated by tissue transglutaminase (tTG). These deamidated gluten-derived peptides are presented to T cells by antigen presenting cells (APCs) via HLA-DQ2 or DQ8.

Searching for susceptibility genes underlying multifactorial disorders

The identification of genes that predispose to multifactorial diseases is hampered by multigenic inheritance and genetic heterogeneity.[57] Multifactorial diseases are caused by variants in multiple genes. Certain combinations of these variants will lead to disease, but the variant by itself is not sufficient to cause the disease. Hence, these variants may occur frequently in the population, and many gene-carriers will not suffer from disease. Furthermore, identical phenotypes can arise that are caused by variants in different genes. This genetic heterogeneity implies that a certain gene variant causing disease in some patients may be absent in other patients. To complicate matters even more, allelic

heterogeneity must also be taken into account. Different variants in the same gene can cause the same phenotype. The HLA-DQ involvement in coeliac disease is a good example of this, since both DQ2 and DQ8 confer risk. All these considerations must be taken into account when searching for susceptibility genes in multifactorial diseases. There are two major approaches for finding these genes: linkage analysis and association studies.[58] Both approaches are widely used and are discussed in more detail below.

Linkage analysis

Linkage analysis is used to localize susceptibility genes within the genome. In general, families with multiple patients are collected, and a genome-wide screen with microsatellite markers, evenly spaced on all chromosomes, is performed. Markers that co-segregate with the disease indicate the regions in the genome that contain disease susceptibility genes. Knowledge about the disease process is not required, which makes linkage analysis applicable for all inherited disorders and traits. For some disorders it is possible to make assumptions about the mode of inheritance of the disease. These disorders can be studied by parametric linkage analysis, in which a genetic model is defined based on disease segregation and recurrence risks in families. Parametric linkage analysis has been very successful in localizing genes causing monogenic disorders. By the end of 2001, 1336 genes had been identified in which mutations cause a Mendelian disorder. In contrast, only seven genes involved in multifactorial diseases have been identified so far.[57] In multifactorial diseases there is no Mendelian inheritance of the disease as multiple genes are involved. Each gene can act in a dominant, recessive or X-linked fashion but its mode of inheritance cannot be recognized in families. It is therefore not possible to define a genetic model for parametric linkage analysis. These disorders can be studied by non-parametric linkage analysis, in which no assumptions are made about the mode of inheritance. This approach is based on allele sharing between affected individuals from one family. Regions with increased sharing of marker alleles identical-by-descent (IBD), when compared to the expected sharing based on the familial relationship between the patients, may contain the disease genes. Susceptibility genes for multifactorial diseases are most often mapped in affected sibling pairs (ASPs) because they are relatively easy to collect. Siblings already share 50% of their DNA on average, so that large numbers of ASPs are needed to detect linkage.[59] For example, approximately 115 ASPs are needed to

detect with 90% probability a locus with a locus-specific λ_s of 2. But to detect loci with small effects, much larger sample sizes are required. In coeliac disease, with a total λ_s of 20, there is a higher probability that some of the loci have moderate or strong effects ($\lambda_s \geq 2$), which reduces the number of ASPs needed to detect them. As well as ASPs, other affected relative pairs can also be used, except for the parent-child relations since they share 50% of their DNA by definition.

Some practical considerations must be taken into account when designing a linkage study. Firstly, families with multiple affected individuals must be collected, which can be difficult. Also, when the patients attend different clinics, homogeneity in diagnosis must be ensured. Secondly, a whole genome screen comprises about 400 microsatellite markers, which must be available. Usually at least 100 ASPs are necessary to give a linkage analysis sufficient power to detect at least one locus, and parents or healthy siblings must also be included to determine whether the alleles are shared IBD. So, with one ASP comprising four individuals that have to be genotyped, at least 160,000 genotypes have to be created for an initial screening of the genome. Many laboratories do not have the resources to carry out these large-scale studies. And thirdly, a genome-wide screen is just the beginning of the search for disease-causing genes. Candidate regions obtained by linkage analysis are usually quite large and contain many genes. The disease-causing gene must be identified by testing variants in the positional candidate genes by association studies.

Association studies

Association analysis is used to test polymorphisms for association with a disease. There are three main applications: testing of functional or positional candidate genes and genome-wide association analysis. Functional candidate genes are selected based on their function and expression profile. This strategy implies knowledge of the disease process and can only be applied to genes with a known function, which includes about one-third of all genes. Positional candidate genes are located within a region of linkage or association. Genes with a function compatible with the disease will have highest priority, but genes with an unknown function also become serious candidates. For whole-genome

association studies no *a priori* knowledge about the disease is needed as SNPs spaced approximately every 10 kb on all chromosomes are tested for association with a disease.

To implicate a gene as disease causing, one or more variants in this gene must be shown to be associated with the disease. The most widely used approach is to test a variant in a number of cases and controls and determine whether allele or genotype frequencies are significantly different between the groups. Independent patients are required for case-control studies, and these can be collected easily, for example via one clinic. However, sometimes the control individuals are harder to collect. Healthy blood donors or samples collected for other purposes are most easily obtained and they are frequently used. It is most important to match the ethnic background of the controls to that of the patients. Allele frequencies can differ between populations and this may lead to false-positive associations. This matter raised strong concerns in the past, but it is now relatively easy to test for the presence of stratification by typing a number of unlinked microsatellite markers.[60] Stratification can be excluded when no significant differences between patients and controls are present.

Large numbers of patients are also needed in association analysis to detect loci with small effects, just as in linkage analysis. Furthermore, when no prior risk alleles are known (for example from other studies or based on the functional characteristics of a certain allele), multiple variants must be tested in one gene. To reduce the number of genotypings, a DNA pooling strategy can be used.[61] Identical amounts of DNA from all patients and controls are pooled together in a patient-pool and a control-pool. Allele frequencies are estimated from these pools and tested for the presence of significant differences. Even when the DNA pools are included in triplicate, together with a few individual samples for correcting PCR artefacts, the total number of genotypings for one variant does not exceed ~25, compared to 400 genotypings for 200 patients and 200 controls for example. Positive associations have to be verified by individual genotyping, as frequencies of some alleles can be over- or underestimated in DNA pools due to PCR artefacts. DNA pooling is particularly useful for large-scale screening studies, in which many microsatellite markers or single nucleotide polymorphisms (SNPs) have to be tested for association. In practice, DNA pooling is applied for whole genome association analysis, refining of candidate regions obtained by linkage analysis, and typing of many SNPs in multiple candidate genes.

The transmission/disequilibrium test (TDT) can be used to avoid effects due to population stratification, but only if both parents of the patients are available.[62] Transmission of variants are scored from heterozygous parents to the patients and compared to a random transmission of 50%, which would be expected if the variant is not associated with the disease. So the non-transmitted alleles serve as control alleles and the possibility of stratification is eliminated. The major problem with TDT is that only heterozygous parents are informative, which greatly reduces the power to detect association. Furthermore, the workload is increased compared to a case-control design, as three individuals must be typed for each patient. Besides, TDT is not suitable for late-onset disorders, as most parents of patients are no longer available. A major advantage of TDT is that phase-known haplotypes of several variants can be determined.

Aims and outline of this thesis

In the past few years, great progress has been made in understanding the pathogenesis of coeliac disease. However, little is yet known about the genes involved in the disease process. For a better understanding of coeliac disease, it is of vital importance to identify and characterize the predisposing genes. The aim of the project described in this thesis was to localize and identify the genes that cause coeliac disease in the Dutch population. Two different approaches were used to achieve this goal: linkage analysis for the localization of the disease genes and association analysis for their identification. Part I of this thesis describes the results of the linkage analysis. A genome-wide screen was performed in Dutch sibpairs affected with coeliac disease and two regions, both conferring a considerable risk to coeliac disease development, were identified (*Chapter 2*). An exceptionally large number of coeliac disease patients was present in one family and the major gene in this family was mapped to a third region (*Chapter 3*). Part II describes the association analysis of candidate regions and candidate genes. The most promising candidate region obtained by the linkage analysis was subjected to systematic fine-mapping using a DNA pooling strategy. Positional candidate genes in this region were subsequently tested for association with coeliac disease (*Chapter 4*). Polymorphisms in three functional candidate genes, located outside regions with linkage to coeliac disease, were also tested for association (*Chapters 5-7*). Finally, an extensive scan of the MHC region was performed in search for additional HLA susceptibility genes, by comparing

DQ2-positive haplotypes from patients to control DQ2 haplotypes (*Chapter 8*). The implications of the results described in this thesis are discussed within the context of the current knowledge of coeliac disease (*Chapter 9*).

References

1. Vogten AJ, Pena AS. Coeliac disease: one century after Samuel Gee (1888). *Neth J Med* 1987; **31:** 253-5.
2. Herter CA. On infantilism from chronic intestinal infection. *Mac Millan Comp NY* 1908.
3. van Berge-Henegouwen GP, Mulder CJ. Pioneer in the gluten free diet: Willem-Karel Dicke 1905-1962, over 50 years of gluten free diet. *Gut* 1993; **34:** 1473-5.
4. Dicke WK. Simple dietary treatment for the syndrome of Gee-Herter. *Neth J Med* 1941; **85:** 1715-6 (in Dutch).
5. Dicke WK. Coeliac disease. Investigation of the harmful effects of certain types of cereal on patients with coeliac disease. (Thesis). University of Utrecht, The Netherlands. 1950**:** (in Dutch).
6. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 1992; **102:** 330-54.
7. Janatuinen EK, Kemppainen TA, Julkunen RJ*, et al.* No harm from five year ingestion of oats in coeliac disease. *Gut* 2002; **50:** 332-5.
8. Rostami K, Kerckhaert J, Tiemessen R, von Blomberg BM, Meijer JW, Mulder CJ. Sensitivity of antiendomysium and antigliadin antibodies in untreated celiac disease: disappointing in clinical practice. *Am J Gastroenterol* 1999; **94:** 888-94.
9. Farrell RJ, Kelly CP. Diagnosis of celiac sprue. *Am J Gastroenterol* 2001; **96:** 3237-46.
10. Dieterich W, Ehnis T, Bauer M*, et al.* Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nat Med* 1997; **3:** 797-801.
11. Tursi A, Brandimarte G, Giorgetti GM. Prevalence of antitissue transglutaminase antibodies in different degrees of intestinal damage in celiac disease. *Clin Gastroenterol* 2003; **36:** 219-21.
12. Ventura A, Magazzu G, Greco L. Duration of exposure to gluten and risk for autoimmune disorders in patients with celiac disease. *Gastroenterology* 1999; **117:** 297-303.
13. Cellier C, Flobert C, Cormier C, Roux C, Schmitz J. Severe osteopenia in symptom-free adults with a childhood diagnosis of coeliac disease. *Lancet* 2000; **355:** 806.
14. Holmes GK, Prior P, Lane MR, Pope D, Allan RN. Malignancy in coeliac disease--effect of a gluten free diet. *Gut* 1989; **30:** 333-8.
15. Corrao G, Corazza GR, Bagnardi V*, et al.* Mortality in patients with coeliac disease and their relatives: a cohort study. *Lancet* 2001; **358:** 356-61.
16. Holmes GK. Screening for coeliac disease in type 1 diabetes. *Arch Dis Child* 2002; **87:** 495-8.
17. Collin P, Kaukinen K, Valimaki M, Salmi J. Endocrinological disorders and celiac disease. *Endocr Rev* 2002; **23:** 464-83.
18. Davison S. Coeliac disease and liver dysfunction. *Arch Dis Child* 2002; **87:** 293-6.
19. Carnicer J, Farre C, Varea V, Vilar P, Moreno J, Artigas J. Prevalence of coeliac disease in Down's syndrome. *Eur J Gastroenterol Hepatol* 2001; **13:** 263-7.
20. George EK, Mearin ML, Bouquet J*, et al.* High frequency of celiac disease in Down syndrome. *J Pediatr* 1996; **128:** 555-7.
21. Jansson U, Johansson C. Down syndrome and celiac disease. *J Pediatr Gastroenterol Nutr* 1995; **21:** 443-5.

22. Mackey J, Treem WR, Worley G, Boney A, Hart P, Kishnani PS. Frequency of celiac disease in individuals with Down syndrome in the United States. *Clin Pediatr (Phila)* 2001; **40:** 249-52.
23. Bonamico M, Pasquino AM, Mariani P*, et al.* Prevalence and clinical picture of celiac disease in Turner syndrome. *J Clin Endocrinol Metab* 2002; **87:** 5495-8.
24. Ivarsson SA, Carlsson A, Bredberg A*, et al.* Prevalence of coeliac disease in Turner syndrome. *Acta Paediatr* 1999; **88:** 933-6.
25. Thapa BR. Celiac disease in India. *Indian J Pediatr* 1999; **66:** S16-20.
26. Demir H, Yuce A, Kocak N, Ozen H, Gurakan F. Celiac disease in Turkish children: presentation of 104 cases. *Pediatr Int* 2000; **42:** 483-7.
27. George EK, Mearin ML, Franken HC, Houwen RH, Hirasing RA, Vandenbroucke JP. Twenty years of childhood coeliac disease in The Netherlands: a rapidly increasing incidence? *Gut* 1997; **40:** 61-6.
28. Csizmadia CG, Mearin ML, von Blomberg BM, Brand R, Verloove-Vanhorick SP. An iceberg of childhood coeliac disease in the Netherlands. *Lancet* 1999; **353:** 813-4.
29. Rostami K, Mulder CJ, Werre JM*, et al.* High prevalence of celiac disease in apparently healthy blood donors suggests a high prevalence of undiagnosed celiac disease in the Dutch population. *Scand J Gastroenterol* 1999; **34:** 276-9.
30. Ivarsson A, Persson LA, Juto P, Peltonen M, Suhr O, Hernell O. High prevalence of undiagnosed coeliac disease in adults: a Swedish population-based study. *J Intern Med* 1999; **245:** 63-8.
31. Hovdenak N, Hovlid E, Aksnes L, Fluge G, Erichsen MM, Eide J. High prevalence of asymptomatic coeliac disease in Norway: a study of blood donors. *Eur J Gastroenterol Hepatol* 1999; **11:** 185-7.
32. Catassi C, Ratsch IM, Fabiani E*, et al.* High prevalence of undiagnosed coeliac disease in 5280 Italian students screened by antigliadin antibodies. *Acta Paediatr* 1995; **84:** 672-6.
33. Hovell CJ, Collett JA, Vautier G*, et al.* High prevalence of coeliac disease in a population-based study from Western Australia: a case for screening? *Med J Aust* 2001; **175:** 247-50.
34. Gomez JC, Selvaggio GS, Viola M*, et al.* Prevalence of celiac disease in Argentina: screening of an adult population in the La Plata area. *Am J Gastroenterol* 2001; **96:** 2700-4.
35. Gandolfi L, Pratesi R, Cordoba JC, Tauil PL, Gasparin M, Catassi C. Prevalence of celiac disease among blood donors in Brazil. *Am J Gastroenterol* 2000; **95:** 689-92.
36. Fasano A, Berti I, Gerarduzzi T*, et al.* Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. *Arch Intern Med* 2003; **163:** 286-92.
37. Mylotte M, Egan-Mitchell B, Fottrell PF, McNicholl B, McCarthy CF. Family studies in coeliac disease. *Q J Med* 1974; **43:** 359-69.
38. Rolles CJ, Myint TO, Sin WK, Anderson M. Proceedings: Family study of coeliac disease. *Gut* 1974; **15:** 827.
39. David TJ, Ajdukiewicz AB. A family study of coeliac disease. *J Med Genet* 1975; **12:** 79-82.
40. Stevens FM, Lloyd R, Egan-Mitchell B*, et al.* Reticulin antibodies in patients with coeliac disease and their relatives. *Gut* 1975; **16:** 598-602.
41. Stokes PL, Ferguson R, Holmes GK, Cooke WT. Familial aspects of coeliac disease. *Q J Med* 1976; **45:** 567-82.
42. Stenhammar L, Brandt A, Wagermark J. A family study of coeliac disease. *Acta Paediatr Scand* 1982; **71:** 625-8.
43. Maki M, Holm K, Lipsanen V*, et al.* Serological markers and HLA genes among healthy first-degree relatives of patients with coeliac disease. *Lancet* 1991; **338:** 1350-3.
44. Vogelsang H, Wyatt J, Penner E, Lochs H. Screening for celiac disease in first-degree relatives of patients with celiac disease by lactulose/mannitol test. *Am J Gastroenterol* 1995; **90:** 1838-42.

45. Greco L, Romino R, Coto I*, et al.* The first large population based twin study of coeliac disease. *Gut* 2002; **50:** 624-8.
46. Sollid LM. Molecular basis of celiac disease. *Annu Rev Immunol* 2000; **18:** 53-81.
47. Sollid LM, Thorsby E. HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 1993; **105:** 910-22.
48. Ivarsson A, Persson LA, Nystrom L*, et al.* Epidemic of coeliac disease in Swedish children. *Acta Paediatr* 2000; **89:** 165-71.
49. Ivarsson A, Hernell O, Stenlund H, Persson LA. Breast-feeding protects against celiac disease. *Am J Clin Nutr* 2002; **75:** 914-21.
50. Ivarsson A, Hernell O, Nystrom L, Persson LA. Children born in the summer have increased risk for coeliac disease. *J Epidemiol Community Health* 2003; **57:** 36-9.
51. Shan L, Molberg O, Parrot I*, et al.* Structural basis for gluten intolerance in celiac sprue. *Science* 2002; **297:** 2275-9.
52. Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* 2002; **2:** 647-55.
53. van de Wal Y, Kooy Y, van Veelen P*, et al.* Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. *J Immunol* 1998; **161:** 1585-8.
54. Molberg O, McAdam SN, Korner R*, et al.* Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat Med* 1998; **4:** 713-7.
55. Molberg O, McAdam SN, Sollid LM. Role of tissue transglutaminase in celiac disease. *J Pediatr Gastroenterol Nutr* 2000; **30:** 232-40.
56. Vader LW, de Ru A, van der Wal Y*, et al.* Specificity of tissue transglutaminase explains cereal toxicity in celiac disease. *J Exp Med* 2002; **195:** 643-9.
57. Glazier AM, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. *Science* 2002; **298:** 2345-9.
58. Borecki IB, Suarez BK. Linkage and association: basic concepts. *Adv Genet* 2001; **42:** 45-66.
59. Risch N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 1990; **46:** 229-41.
60. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999; **65:** 220-8.
61. Sham P, Bader JS, Craig I, O'Donovan M, Owen M. DNA Pooling: a tool for large-scale association studies. *Nat Rev Genet* 2002; **3:** 862-71.
62. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52:** 506-16.

Part I

Linkage analysis

Chapter 2

A genome-wide screen in affected sibpairs

A major non-HLA locus in coeliac disease maps to chromosome 19

M.J. van Belzen, J.W.R. Meijer, L.A. Sandkuijl, A.F.J. Bardoel, C.J.J. Mulder, P.L. Pearson, R.H.J. Houwen and C. Wijmenga

Gastroenterology 2003; 125: 1032-1041.

Summary

Background & Aims: The pathogenesis of coeliac disease is still unknown, despite its well-known association with HLA-DQ2 and DQ8. It is clear that non-HLA genes contribute to coeliac disease development as well, but none of the previous genome-wide screens in coeliac disease have resulted in identification of these genes. *Methods:* We therefore performed a two-stage genome-wide screen in 101 affected sibpairs from 82 Dutch families that met strict diagnostic criteria. The small intestinal biopsy samples, on which the original coeliac disease diagnoses had been based, showed a Marsh III lesion in all patients on re-evaluation by one pathologist. For association analysis of markers in regions linked to coeliac disease, 216 independent MIII patients and 216 age and sex-matched controls were available. *Results:* As expected, highly significant linkage to the HLA-region was detected (multipoint maximum lod score (MMLS) = 8.14). More importantly, significant linkage was also present at 19p13.1 (MMLS = 4.31), with the peak at marker D19S899. Moreover, this marker was also significantly associated with coeliac disease in the case-control study (corrected p = 0.016). Furthermore, we identified suggestive linkage to 6q21-22, which is ~70 cM downstream from the HLA-region (MMLS = 3.10). *Conclusions:* Significant linkage of coeliac disease to chromosome region 19p13.1 was detected in our genome-wide screen. These results were confirmed by the association of D19S899 to coeliac disease in an independent case-control cohort. Furthermore, we identified a possible second coeliac disease locus on chromosome region 6q21-22.

Introduction

Coeliac disease is one of the most common forms of food intolerance in the Western world, with an estimated prevalence as high as 0.3-1%.[1] Coeliac disease is considered to be an autoimmune disorder, and it is strongly associated with the Human Leukocyte Antigen (HLA) region. Coeliac disease patients are unable to tolerate gluten from wheat, barley and rye. The gluten peptides that arrive in the small intestine of a coeliac disease patient are presented to T-cells and this process leads to the characteristic coeliac disease lesion in the gut, with villous atrophy, hyperplastic crypts and large infiltrates of lymphocytes into the epithelium and the lamina propria.[2] The presence of the small intestinal lesion can lead to a broad variety of symptoms, varying from diarrhoea and

abdominal distension to fatigue, anaemia, osteoporosis and short stature due to inefficient uptake of nutrients from food.[1]

Coeliac disease is a strongly inheritable disorder, with a ~10% recurrence risk for siblings of a patient.[3] Recently, it was shown that the disease concordance rates are 86% in monozygous twins and 20% in dizygous twins, indicating a strong genetic contribution to coeliac disease development.[4] The association between coeliac disease and the HLA-region is well known. Over 90% of coeliac disease patients express the HLA-DQ2 protein, encoded by the DQA1*0501 and DQB1*02 alleles. The majority of the DQ2-negative patients carry the DQA1*0301 and DQB1*0302 alleles, which combine to form the DQ8 molecule.[5] HLA-DQ2 and DQ8 have been shown to bind gluten-derived peptides, which provokes a T-cell response, thereby explaining the genetic association.[3,6] However, the genetic contribution of coeliac disease cannot be explained by HLA solely. The concordance rate in monozygous twins is much higher than in HLA-identical siblings (~30%), indicating the action of non-HLA genes. This is compatible with the fact that the prevalence of DQ2 in the general population is ~25%, which is much higher than the prevalence of coeliac disease.[5]

Since coeliac disease is the only autoimmune disorder in which both an important genetic (HLA-DQ) and environmental (gluten) factor are known, this disease presents a unique paradigm for studying complex autoimmune diseases. Elucidation of the genes involved in coeliac disease pathogenesis will therefore not only contribute to our knowledge about coeliac disease pathogenesis, but may also provide a beginning to understanding the destructive processes in other autoimmune disorders. So far, eight whole genome screens have been performed to identify the non-HLA genes causing coeliac disease. Unfortunately, none were able to identify genome-wide statistically significant linkage outside the HLA-region except in a population isolate[7] and chromosomal regions showing evidence for linkage differed markedly between the studies.

We set out to perform a genome-wide scan in Dutch sibpairs affected with coeliac disease that met strict diagnostic criteria. We attempted to retrieve the small intestinal biopsies of all patients initially selected for this study. Only those sibpairs whose biopsies could be re-examined and which showed the characteristic coeliac disease lesion on re-evaluation by one experienced pathologist were included in this genome-wide screen.

This strategy is expected to lead to a smaller, but more homogeneous set of affected sibpairs, thereby increasing the chance of detecting significant linkage.

Materials and Methods

Study subjects

Families with two or more siblings affected with coeliac disease were invited to participate in our study through an advertisement in the Dutch Coeliac Society's newsletter. Several families were also referred by their medical specialists. Blood was collected from those families in which the affected siblings were diagnosed by a small bowel biopsy. All biopsy samples on which the original coeliac disease diagnosis had been established were collected and sent to an experienced pathologist (JWRM) for re-evaluation and Marsh classification. This procedure ensured homogeneity in the evaluation of the biopsy samples, since the patients were originally diagnosed in more than 50 different hospitals over the past 30 years. Only families with at least two siblings with coeliac disease grade Marsh IIIa, Marsh IIIb or Marsh IIIc (i.e. partial, subtotal or total villous atrophy respectively, with the presence of crypt hyperplasia and an increased number of intraepithelial lymphocytes (>30 per 100 enterocytes))[8] were included in our genome-wide screen. Patients with Marsh II or Marsh I biopsy samples, or whose biopsy samples were too damaged to classify or unavailable were excluded from this study. This procedure resulted in exclusion of 23 affected sibpairs from the study. In total, 67 families with two or more Marsh III siblings were included in the whole genome screen. They comprised 60 families with two affected siblings, six families with three affected siblings and one family with four affected siblings. Additional siblings from families with missing parents were included to determine the parental genotypes. The composition of the families, who came from all parts of the Netherlands and were of Dutch origin, is shown in Table 1.

For the follow-up study of the chromosome regions showing evidence for linkage, 15 more families were recruited, 14 with two affected siblings and one with three affected siblings. These families met the same criteria as the first-stage families, i.e. all affected siblings had a Marsh III lesion. The composition of the families is also shown in Table 1. Thirteen of these families were of Dutch origin and the other two were from the Dutch-speaking part of Belgium.

Table 1. Composition of the families with 2 or more siblings affected with coeliac disease.

Stage	Family type	Number
Initial genome-wide screen	2 parents	43
(N = 67)	1 parent	
	+ *1 additional sibling*	9
	+ *0 additional siblings*	2
	0 parents	
	+ *2 additional siblings*	11
	+ *1 additional sibling*	1
	+ *0 additional siblings*	1
Follow-up	2 parents	12
(N = 15)	1 parent	
	+ *1 additional sibling*	1
	0 parents	
	+ *2 additional siblings*	2
	Total	82

In total, 74 families with two affected siblings, seven families with three affected siblings and one family with four affected siblings were included in this study. This resulted in 101 affected sibpairs, consisting of both children as well as adults. The affected siblings were typed at the HLA-DQA1 and DQB1 loci by SSCP based methods.[9] The patients could be divided into two groups based on their age at the time of diagnosis. The childhood-onset group consisted of 47 families with 55 affected sibpairs who were diagnosed before the age of 11. The adult-onset group consisted of 33 families with 42 affected sibpairs who were diagnosed after the age of 20. Two families fell between these categories.

For association analysis in the regions with evidence for linkage, 216 independent Dutch coeliac disease patients were available. Of these patients, 141 were female (65%) and 75 were male (35%). The age ranged from 4-91 years with a mean age of 39 years. These patients met the same strict diagnostic criteria as those selected for the genome-wide screen; all showed a Marsh III lesion upon re-evaluation of their initial biopsy specimens by the same pathologist (JWRM). Control individuals were age and sex-matched to the coeliac disease patients and consisted of random hospital controls. The study was approved by the Medical Ethics Committee of the University Medical Centre Utrecht and written informed consent was obtained from all the participants.

Genotyping the microsatellite markers

A total of 326 microsatellite markers were genotyped in the 67 families in the initial genome-wide screen. Genotypes were obtained for 270 subjects, including the parents of the affected siblings to gain more information about the number of alleles shared identical by descent at the marker loci. If the parents were unavailable, a maximum of two additional siblings were included to reconstruct the missing parental genotypes. Our marker set consisted mainly of screening set 6 from the Marshfield Centre for Medical Genetics, complemented with markers selected from the Marshfield genetic map (http://research.marshfieldclinic.org/genetics/). The markers were evenly spread over all 22 autosomes and the X chromosome, with an average spacing of 12 cM. A denser marker map was used for the *CELIAC1* region (HLA), the *CELIAC2* region on chromosome 5q31-q33 and the *CELIAC3* region on chromosome 2 (CTLA4/CD28), since these regions had been previously implicated as candidate regions in other genome scans.[10-16]

Forty-three additional markers were typed in six non-HLA regions that showed evidence for linkage in the initial genome-wide screen at a nominal significance level ≤0.01 (lod score >1.3). The additional 15 families were also typed for these regions, as well as for the *CELIAC1* and *CELIAC2* regions. The *CELIAC3* region was not typed in these families because no evidence for linkage to this region was present in the initial genome-wide screen.

The markers were amplified in multiplex PCR reactions in 96-well plates. Each plate contained DNA from the coeliac disease families, a negative control, three CEPH reference samples (1331-01, 1331-02 and 1347-02) and four blind control samples that were duplicates of samples in one of the other plates. The reaction volume of 10 µl contained 25 ng of DNA, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 50 ng fluorescence-labelled primer and 0.4 U AmpiTaq Gold (PE Applied Biosystems, Foster City). The PCR products were pooled and separated on a 3700 DNA sequencer (PE Applied Biosystems) and analysed by Genescan 3.5 and Genotyper 2.0 software (PE Applied Biosystems). All genotypes were checked independently by two researchers (MJvB and AFJB).

Statistical analysis

The power to detect linkage to a coeliac disease susceptibility locus in our data set was estimated according to Risch.[17] Probabilities were calculated for detecting loci with different sibling relative risks (λ_s): 1.5, 2, 2.5 and 3. For these calculations we aimed at a p-value of 0.0007, which corresponds to the genome-wide threshold for suggestive linkage.[18]

The genotype data of the genome-wide screen were analysed using the MAPMAKER/SIBS program,[19] which is based on allele sharing between the affected siblings. This program performs a non-parametric linkage analysis, which is preferred since the mode of inheritance of the coeliac disease susceptibility genes is unknown. The identical by descent allele sharing between pairs of affected siblings (0, 1 or 2 alleles) is determined at the marker loci and the observed frequencies are compared to the expected allele sharing of 0.25, 0.5 and 0.25 identical by descent under the null hypothesis of no linkage. Single and multipoint analyses were performed, both allowing for dominance variance (possible triangle method). Multiple affected sibpairs from the same family were treated as independent sibpairs, by calculating the number of alleles shared identical by descent of all possible pairs in the unweighted mode. Allele frequencies were calculated from the data set, since population frequencies of the markers were unknown in our population. The order and location of the markers on the chromosomes was based on the Marshfield genetic map (http://research.marshfieldclinic.org/genetics/).

Several quality checks were performed on the markers before inclusion in the linkage analysis. Firstly, the genotypes of the CEPH reference samples were compared to the CEPH genotype database (http://www.cephb.fr/) to check that the intended marker was amplified. Secondly, the blind controls were checked by an independent investigator and markers with more than one discrepancy were re-analysed. The marker was excluded from the linkage analysis if the discrepancies remained. Thirdly, a Mendelian inheritance check was performed and markers with more than two Mendelian errors were also excluded from the linkage analysis to ensure the quality of the marker set. New alleles are frequently observed in microsatellite markers, since they mutate easily. Therefore, when a new allele was observed in one or two families, only those families were excluded for that marker, but it was included in the other families. In contrast, markers with null-alleles were excluded in all families. A null-allele is an allele that is segregating in a family, but it

is not amplified by PCR due to polymorphisms in the primer-binding site. This can be observed in families in which non-paternity has been excluded, but where one parent is homozygous for a certain allele and a child is homozygous for another allele transmitted by the other parent. Fifteen markers were excluded due to these quality checks, resulting in 311 markers that were used for the linkage analysis.

The initial genome-wide screen was analysed in the 67 families typed for all 311 markers. The seven regions showing evidence for linkage at nominal significance ≤0.01, corresponding to a lod score of 1.3,[18,20] were further analysed in all 82 families. P-values corresponding to the obtained lod scores were approximated according to Holmans.[20] The threshold levels for linkage were according to the genome-wide significance levels as proposed by Lander and Kruglyak,[18] with suggestive linkage at $p = 7 \times 10^{-4}$, significant linkage at $p = 2 \times 10^{-5}$ and highly significant linkage at $p = 3 \times 10^{-7}$ (corresponding lod scores are 2.6, 4.0 and 5.8, respectively).

One could hypothesize that different genes are involved in childhood-onset and adult-onset coeliac disease. We tested this hypothesis in the two regions with evidence for linkage, 6q21-22 and 19p13.1. The families were divided into one group with 47 families with childhood-onset coeliac disease and another group with 33 families with adult-onset coeliac disease (see patients section). The linkage analysis was repeated in both age groups. The two families that did not fit into these categories were not included in this analysis.

Markers from the genome-wide screen in regions with evidence for linkage were tested for association in a case-control study. Eight markers from the 6q21-22 region (D6S1563 to D6S1712, see Table 2) and five markers from the 19p13.1 region (D19S714 to D19S215) were typed in 216 Marsh III coeliac disease patients and 216 age- and sex-matched controls. Overall significance of the markers was obtained by means of an Nx2 chi-square test (N = number of alleles at the marker locus with a frequency ≥ 0.05, plus one group of the remaining alleles). A Bonferroni correction was applied to correct for the thirteen markers that were tested. Association of marker D19S899 with coeliac disease was tested in the 82 families from the linkage analysis using the sib_tdt implementation of the ASPEX package. This program calculates empirical probabilities for association independent of linkage within families.

Results

Initial genome-wide screen

Sixty-seven families with 84 affected sibpairs were available for the initial genome-wide screen. Power calculations were performed to estimate the chance to detect susceptibility loci with different relative risks to a sibling of a coeliac disease patient (λ_s). It was estimated that with this set of families the chance to detect a locus with a λ_s of 1.5 was 20%, a λ_s of 2.0 was 72%, a λ_s of 2.5 was 94% and λ_s of 3.0 was 99%. A whole genome screen with microsatellite markers was performed in these patients and their parents. The multipoint maximum lod scores (MMLS) on all chromosomes are depicted in Figure 1.

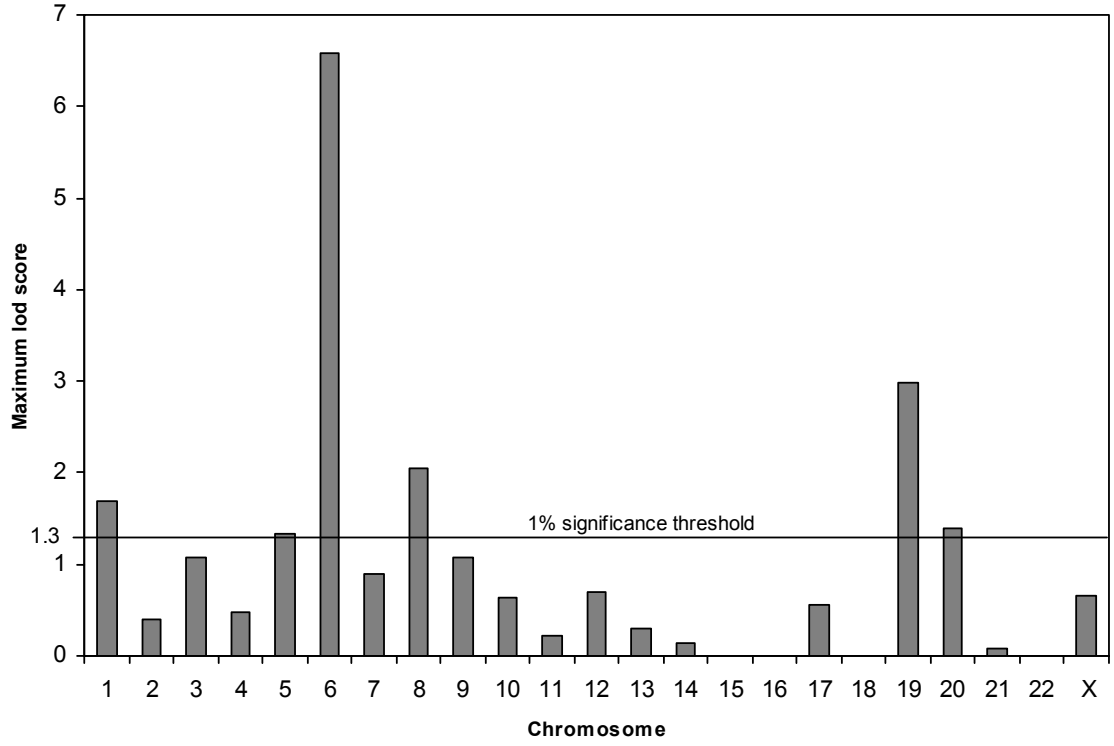


Figure 1. Maximum lod scores from the multipoint linkage analysis present on each chromosome in the initial genome-wide screen in 67 families. The horizontal line indicates a nominal significance threshold of 0.01. All regions reaching this threshold were further investigated in a follow-up study.

Genome-wide significant linkage to the *CELIAC1* locus in the HLA-region on chromosome 6 was detected (MMLS = 6.58). Six other regions, on chromosomes 1p, 5q, 6q, 8q, 19 and 20p, were identified with nominal p-values ≤0.01 (MMLS > 1.3). The threshold for suggestive linkage was reached at chromosome 19 (MMLS = 2.98). The

region on chromosome 6q is approximately 70 cM downstream from the HLA-region and yielded an MMLS value of 1.92.

Follow-up analysis of chromosomes 1, 5, 6, 8, 19 and 20

The 15 additional families were typed for the markers from the seven regions with an MMLS >1.3. Furthermore, a total of 43 new markers were added to the interesting regions on chromosomes 1p, 6q, 8q, 19 and 20p to obtain a spacing of <5 cM, and these were typed in all 82 families. No extra markers were typed in *CELIAC1* and *CELIAC2* regions (HLA and 5qter, respectively), as these regions were already densely covered with markers in the initial genome-wide screen. The results of linkage analysis of chromosomes 1, 5, 6, 9, 19 and 20 after fine-mapping are shown in Figure 2. For completeness, the linkage graphs of the other chromosomes after the initial genome-wide screen were also included in this figure.

Suggestive linkage to chromosome 19 was detected in the initial genome-wide screen, between D19S714 on the p-arm and D19S587 on the q-arm. Twelve new markers were added to this region, leading to a map with a 1.0-4.3 cM spacing. This resulted in an increase of the MMLS value from 2.98 to 4.43 (nominal $p = 6.2 \times 10^{-6}$) at marker D19S899, reaching genome-wide significance. Genotyping all 82 families resulted in a comparable MMLS of 4.31 at the same position. This novel coeliac disease locus is located at 19p13.1.

Genotyping twelve extra markers in the chromosome 6q region resulted in an increase of the MMLS value from 1.92 to 3.14 (nominal $p = 1.3 \times 10^{-4}$) in the initial 67 families at region 6q21-22. Typing all 82 families for these markers resulted in a similar MMLS of 3.10. The MMLS at this locus peaked at marker D6S1608 at 122 cM and reached the threshold for suggestive linkage.

Adding more markers to chromosomes 1p, 8q and 20p did not increase evidence for linkage in the initial 67 families. Genotyping these markers in all families did not change the results for chromosomes 1 and 20. However, the MMLS value on chromosome 8 dropped considerably from 1.96 to 0.75 after genotyping all families.

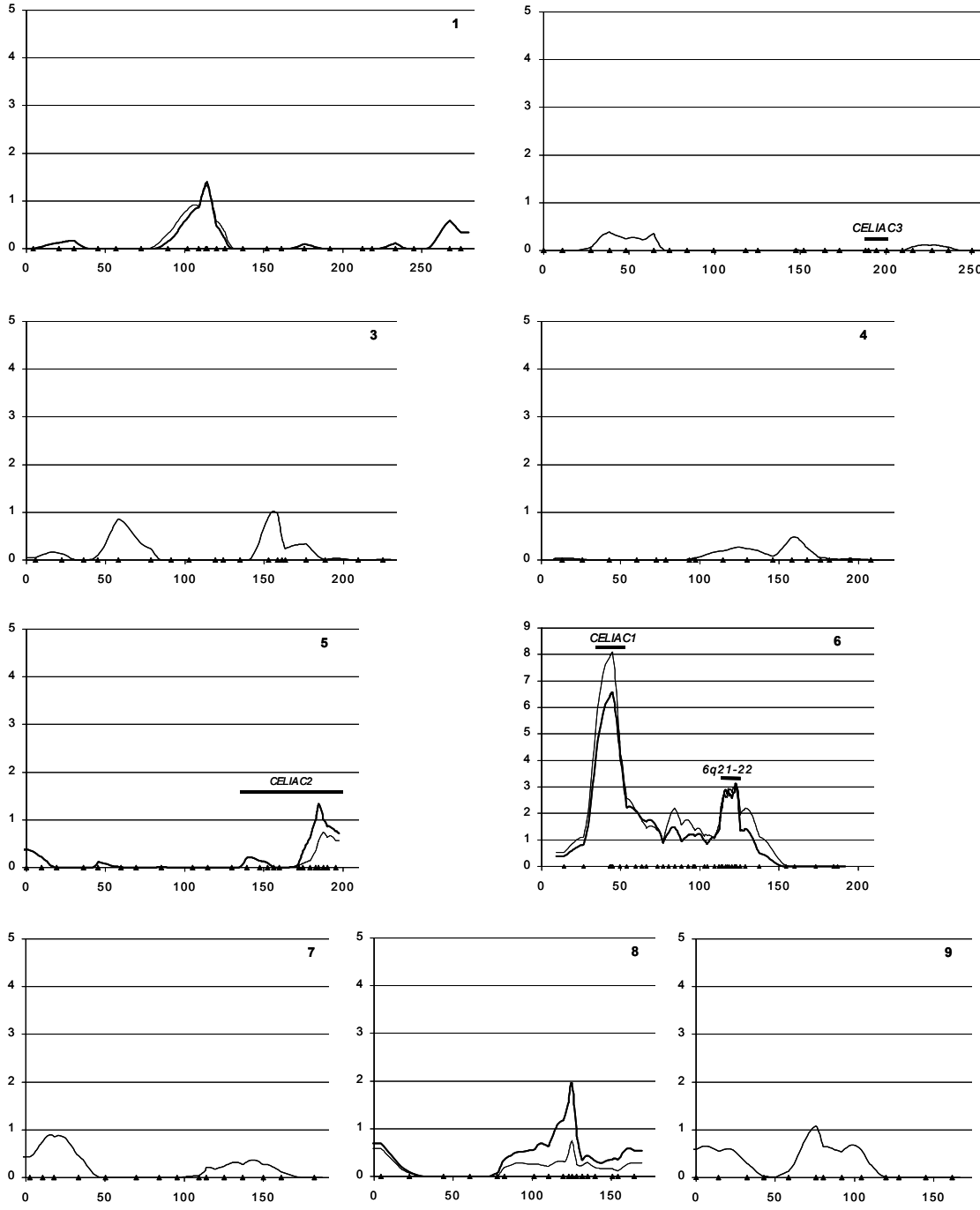
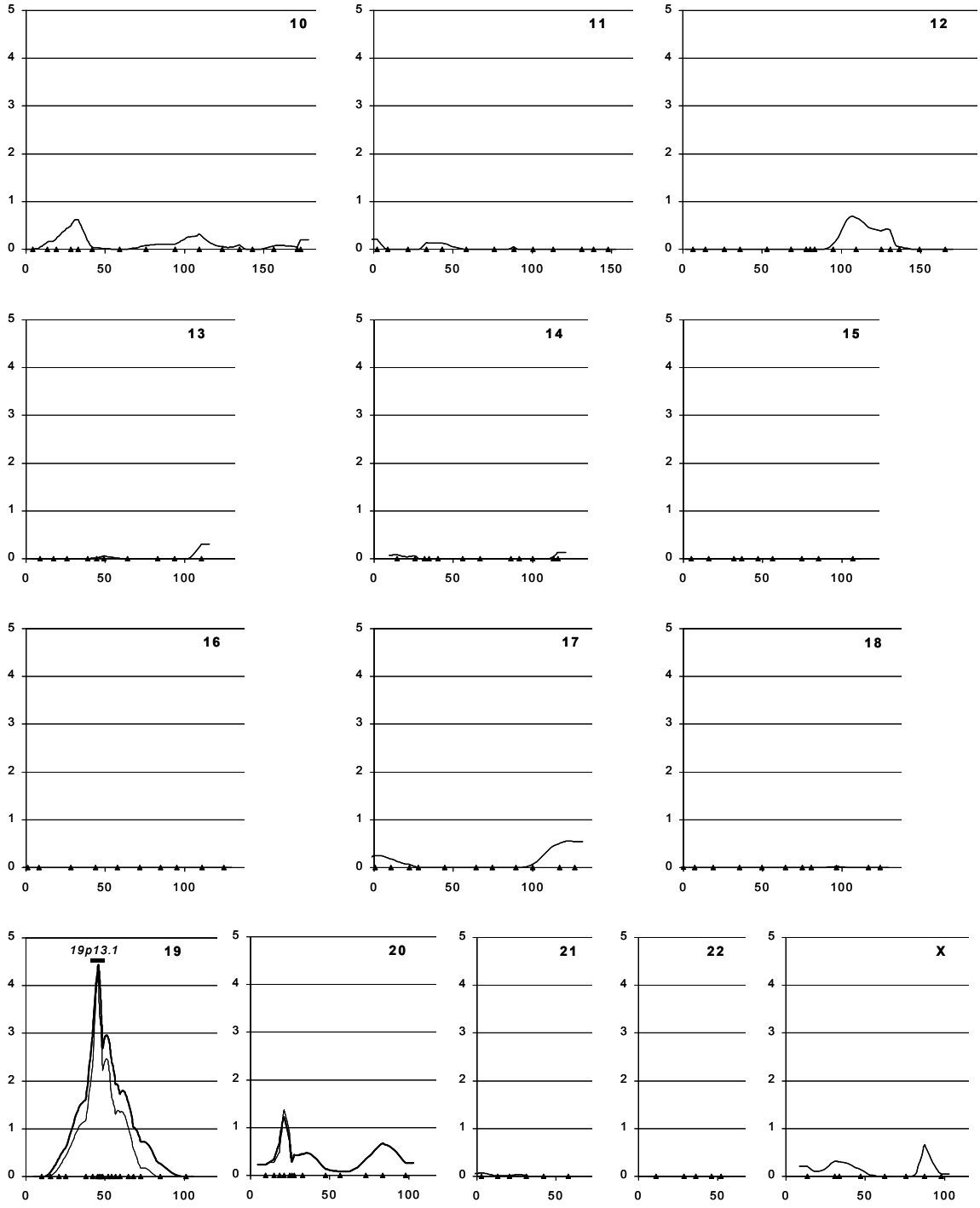


Figure 2. Distribution of MLS values in the initial 67 families (bold line) and in all 82 families (thin line). Graphs of chromosomes 1, 5, 6, 8, 19 and 20 are after fine-mapping of regions with a lod score >1.3; graphs of the other chromosomes represent the results of the initial genome-wide screen. The triangular symbols denote the positions of the markers on the chromosomes.


10
11
12
13
14
15
16
17
18
19p13.1
19
20
21
22
X

Typing the *CELIAC1* region in all 82 families increased the MMLS value from 6.58 to 8.14 (nominal $p = 1.1 \times 10^{-9}$), but the MMLS value in the *CELIAC2* region on chromosome 5 dropped from 1.34 to 0.74. There was no evidence for linkage to the *CELIAC3* region on chromosome 2 (MMLS = 0.0).

The contributions of the19p13.1 and 6q21-22 loci to familial clustering of coeliac disease were determined as proposed by Risch[21]. The relative risk for a sibling due to a certain locus ($\lambda_{s,\ locus}$) equals the prior probability of affected siblings sharing 0 alleles (0.25) divided by the observed probability of no allele sharing. This resulted in a λ_s of 2.6 for the 19p13.1 locus and a λ_s of 2.3 for the 6q21-22 locus, compared to a λ_s of 4.6 for the HLA region.

HLA-DQ typing

In 78 of the 82 families, all affected siblings were HLA-DQ2 positive. In three of the remaining families, at least one of the affected siblings carried HLA-DQ2, while the other patients were HLA-DQ8 positive. In the only DQ2-negative family, both affected siblings were HLA-DQ8 positive.

Stratification for age of diagnosis

The families were divided in two age groups depending on the age of diagnosis to determine whether the 6q21-22 and 19p13.1 are specifically involved in childhood- or adult-onset coeliac disease. Linkage analysis of the two regions was performed in both age groups separately. There was no difference in linkage to these regions between the families with childhood-onset coeliac disease and the families with adult onset coeliac disease, both age-groups contributed almost equally to the lod score (data not shown).

Single point linkage analysis in the putative non-HLA regions

A single point linkage analysis was also performed in the non-HLA the regions that showed moderate evidence for linkage (MMLS > 1.3, corresponding to a nominal p-value ≤0.01) to determine the support for linkage to these regions (Table 2). Ten consecutive markers on chromosome 19 in a 22 cM interval yield lod scores >1 in the initial 67 families, indicating the importance of this region. Interestingly, the lod score of marker

Table 2. Single point lod scores for markers in non-HLA regions with a maximum lod score >1.3 in the multipoint analysis of the initial genome-wide screen.

Chromosome	Marker	cM[a]	Mb[b]	lod score	
				Initial 67 families	All 82 families
1	D1S1728	109.04	82.17	0.22	0.30
	D1S551	113.69	83.07	1.73	1.70
	D1S2807	114.24	83.49	0.33	0.44
5	D5S211	182.89	173.51	0.40	0.08
	D5S498	184.66	174.07	1.14	0.41
	D5S2111	187.81	-	1.14	0.45
	D5S2008	190.18	177.72	0.37	0.25
	D5S408	195.49	180.12	0.30	0.02
	D5S2006	197.54	180.48	0.83	1.26
6	D6S283	109.19	102.30	1.89	2.24
	D6S1021	112.20	104.51	1.50	1.64
	D6S1563	113.60	105.50	1.31	1.76
	D6S278	116.26	108.16	2.74	1.78
	D6S1594	117.29	108.34	1.21	1.31
	D6S474	118.64	112.71	1.73	2.01
	D6S261	120.31	133.99	1.23	1.43
	D6S433	121.97	118.48	2.39	2.38
	D6S1608	122.51	120.68	1.61	1.11
	D6S1712	122.51	121.90	2.43	2.25
8	GATA123H10	122.96	-	0.25	0.01
	D8S592	125.27	117.44	2.28	1.44
	D8S198	128.16	122.44	0.57	0.19
19	D19S714	42.28	16.11	1.47	1.08
	D19S899	45.48	17.73	3.92	3.81
	D19S460	-	18.87	2.14	1.60
	D19S407	48.14	20.56	4.83	3.95
	D19S215	48.52	22.12	1.88	1.23
	D19S433	51.88	31.03	2.17	1.78
	D19S414	54.01	32.53	1.39	1.05
	D19S868	56.69	-	1.07	0.67
	D19S587	59.36	35.83	1.11	0.89
	D19S400	64.70	42.17	1.38	0.90
20	D20S194	18.26	6.13	0.22	0.15
	D20S907	21.15	-	1.49	1.58
	D20S917	24.70	9.28	0.21	0.47

[a] cM positions are based on the genetic map of the Marshfield database.
[b] Physical position according to the December 2002 release from the Ensembl database.

D19S407 at position 48.14 cM is 4.83, which is even higher than its result in the multipoint analysis. Seven markers with lod scores >1 were detected in all 82 families. In the chromosome 6q21-22 region, 10 consecutive markers, spanning 13 cM, also yielded lod scores >1. In the initial screen, two consecutive markers on chromosome 5qter with a lod score of 1.14 were detected, D5S498 and D5S2111. On analysing all 82 families, the lod scores for these markers dropped below 0.5, but the most telomeric marker (D5S2006) increased to 1.25. On chromosomes 1, 8 and 20, only one marker showed linkage at $p \leq 0.01$, while the flanking markers gave no supporting evidence. The positive results on these chromosomes in the multipoint analyses were probably due to these markers solely. The physical locations of the markers were determined to check their order (Table 2). Unfortunately, four markers could not be mapped to the human genome sequence (December 2002 release of the Ensembl database). All the remaining 31 markers could be mapped and were in the correct order. Marker D19S460 was not present on the Marshfield genetic map.

Association analysis

The allele frequencies of the markers in the 6q21-22 and 19p13.1 regions were determined in 216 independent Marsh III patients and 216 control individuals. A significant difference between the cases and controls was detected for marker D19S899, the marker at which the multipoint lod score peaked. For this marker 214 cases and 210 controls could be genotyped, yielding an overall p-value of 0.0013. This result was still significant after correcting for the 13 markers that were tested ($p = 0.016$). The association of D19S899 was not due to one allele, but four out of the six tested alleles were increased in cases. Therefore, lack of power prohibited us to confirm this association in the 82 families used for the linkage analysis. None of the other markers showed significant association with coeliac disease in the case-control cohort.

Discussion

Sixty-seven families with affected sibpairs of Dutch ancestry were available for the initial genome-wide screen. A total of 84 affected sibpairs were present in these families. It was shown by power calculations that this data set had a very high chance of detecting a coeliac disease susceptibility locus with a locus-specific $\lambda_s \geq 2.5$ (> 94%). Also, the chance

to detect a locus with a λ_s of 2.0 is still reasonably good (72%). However, this data set has limited power to detect loci with smaller effects, as is shown by the 20% chance to detect a locus with a λ_s of 1.5. Therefore, this data set has sufficient power to detect loci with moderate to high effects, but loci with very small effects may be missed. The total estimated λ_s for coeliac disease is 25, based on a 10% sibling recurrence risk and a population prevalence of 0.4%. Considering an estimated λ_s of 3-5 for the HLA-region,[21-23] the majority of the genetic susceptibility is due to unknown genes. This data set has a good chance to detect at least one of these loci.

Multipoint linkage analysis revealed the presence of two non-HLA loci predisposing to coeliac disease in the Dutch population. A major non-HLA locus is located on chromosome 19. This region showed genome-wide significant linkage, with an MMLS value of 4.43 (nominal $p = 6.2 \times 10^{-6}$). This finding was supported by single point analysis, showing ten successive markers with lod scores >1 in this region. The results in all 82 families were in agreement with these findings. This high-risk coeliac disease locus accounts for a 2.6-fold increased risk to siblings, which is about half of the risk attributed by the HLA-region in our families ($\lambda_{s,\,HLA} = 4.6$). It is involved in development of coeliac disease in children as well as in adults. Moreover, marker D19S899 was associated with coeliac disease in an independent case-control cohort. The maximum lod score in the linkage analysis was obtained at the same marker, indicating that a coeliac disease susceptibility gene is located near this marker. Importantly, the case-control study revealed that four out of six tested D19S899 alleles were over-represented in the cases compared to the controls. These results imply either the presence of multiple independent disease-causing variants that, in turn, are present on different haplotypes, or a single disease-causing haplotype that shows association with different D19S899 alleles. The observed association with multiple alleles may also explain the lack of power to replicate these findings in the 82 families from the linkage analysis, since only a subset of the families will show association for each one of these four associated alleles. Altogether, the presence of association in the 19p13.1 region, which has been obtained in a different patient sample and by a different approach, provides independent evidence for the presence of a coeliac disease gene in this region.

The linkage graph on chromosome 19 peaked at 19p13.12, although the region that may contain the coeliac disease gene could be much larger. Ten markers with lod scores >1 were present in the initial 67 families, and seven in all 82 families, resulting in a candidate region of at least 11-22 cM (16-26 Mb) ranging from 19p13.12 to 19q13.2. For identification of this susceptibility gene, one may start testing genes in the support interval defined by the maximum lod score –1. This region spans from approximately 43.5 and 47 cM (3Mb) and contains 92 known and predicted genes, according to the Ensembl database. It can however not be excluded that the gene is located outside this interval.

Stratification of the families into two groups, HLA-DQ2 positive and HLA-DQ2 negative, and re-analysing the data in these groups would in theory distinguish the non-HLA loci that interact with HLA-DQ2 from those that act independently. Unfortunately, this is not possible in our data set, since HLA-DQ2 is present in all but one family.

So far, eight other genome-wide scans, all in Caucasian patients, have been undertaken to identify coeliac disease susceptibility loci. One of these studies included patients from a population isolate, whereas the others were undertaken in more outbred populations (Table 3). Three of these studies showed some evidence for linkage to chromosome 19, but at clearly different loci. Weak evidence for linkage was detected twice at 19p13.3 at ~10-16 cM, with an MMLS of 0.84 (nominal p = 0.0544)[16] and a non-parametric lod score (NPL) of ~2.1 (nominal p = 0.02)[24] respectively. At chromosome region 19q13.4, an MMLS value of 1.84 (nominal p = 0.0544)[16] was reported at 92 cM and a recessive heterogeneity lod score (HLOD) of 1.6 at the same position.[25] Thus, the locus on 19p13.1 seems to be unique to the Dutch coeliac disease population. Interestingly, linkage of inflammatory bowel disease (IBD) to this region has been reported,[26] perhaps representing a shared genetic background for intestinal inflammation.

Suggestive linkage to chromosome 6q21-22 was detected in our genome-wide screen, with an MMLS value of 3.14 (nominal p = 1.3 x 10^{-4}). The region of interest spans from approximately 113 to 125 cM, which is clearly distinct from the HLA-region at ~45 cM. Single point analysis revealed that the linkage in this region was contributed by ten consecutive markers, providing support for the presence of another coeliac disease gene on chromosome 6. The λ_s of this locus is with 2.3 comparable to that of the 19p13.1 locus. This locus is also involved in both childhood- and adult-onset coeliac

Table 3. Overview of linkage to non-HLA regions in genome-wide screens conducted for coeliac disease.

Population	No. of families ($1^{st}/2^{nd}/3^{rd}$)[a]	Study design	Suggestive[b] linkage	Significant[b] linkage	Reference
West Ireland	15/-	Affected sibpairs	6p23, 11p11	-	16
Italy	39/57/87	Affected sibpairs	5qter	-	10,11
UK	16/34	Extended families	-	-	25,33
Sweden/ Norway	70/36	Affected sibpairs	-	-	13
Finland	60/38	Affected sibpairs	4p15	-	12
Finland	9/1	Population Isolate	-	15q12	7
North Europe	24/-	Extended families	-	-	24
North America	62/-	Extended families	3p26, 5p14, 18q23	-	34
Netherlands	67/15	Affected sibpairs	6q21-22	19p13.1	this study

[a] Number of families in initial genome-wide screen (1^{st}) and subsequent follow-up studies (2^{nd} and 3^{rd}).
[b] Suggestive linkage: $p < 7.4 \times 10^{-4}$; significant linkage: $p < 2.2 \times 10^{-5}$ (according to criteria proposed by Lander and Kruglyak[18]). The "-" symbol denotes that the indicated genome-wide significance threshold was not reached.

disease. Evidence for linkage of coeliac disease to region 6p12 has been found previously. At 74 cM, a maximum lod score of 2.2 (nominal p = 0.001) and an MMLS value of 1.42 (nominal p = 0.0186) were identified.[16,25] Since our linked region is 40-50 cM apart from the 6p12 locus, it probably represents a distinct locus on chromosome 6. Linkage of type 1 diabetes mellitus to the 6q21 region has been reported several times.[27] Furthermore, a locus around 100 cM (6p15-16.3) has been implicated in systemic lupus erythematosus[28] and in psoriasis.[29] Taken together, these results indicate the possibility of a gene in the chromosome 6q15-22 region that is involved in the autoimmune process.

Positive findings at a nominal p value ≤0.01 were present in four other regions in our initial genome-wide screen: 1p31.1, 5q35.2, 8q24.11 and 20p12.3. Single point analysis of the markers on chromosomes 1, 8 and 20 revealed that the positive results in the multipoint analysis were due to only one marker and there was no support from the flanking markers, suggesting that these loci are false-positive linkage results. This is in agreement with the number of false-positive findings that can be expected in a genome-

wide screen. Simulation of a whole genome screen in 100 affected sibpairs, with no trait locus segregating in the families, showed 22 regions reaching nominal significance at p ≤0.05, three of those reaching a significance level of ≤0.01 and one even reaching the threshold for suggestive linkage.[18] Importantly, no significant linkage was present in this simulated genome-wide screen.

It is difficult to interpret our findings in the *CELIAC2* region on chromosome region 5qter. Some evidence for linkage to this region was detected in the initial 67 families, but the addition of 15 more families resulted in a decrease of the MMLS value from 1.34 to 0.74. This might indicate heterogeneity in our families, but the possibility of a false-positive result cannot be excluded. Suggestive linkage to the *CELIAC2* region was identified in Italian affected sibpairs,[10,11] and support for linkage to this region was present in three other populations,[12,13,16] indicating the presence of a minor coeliac disease susceptibility gene in this region. The CTLA4/CD28 region on chromosome 2 (*CELIAC3*) and the chromosome 11p11 region have been implicated in coeliac disease in previous studies, but our results do not support linkage to these regions in the Dutch population.

In conclusion, two novel coeliac disease loci were identified in the Dutch population. Evidence for linkage to 19p13.1 reached the threshold for genome-wide significance. This locus is the major non-HLA coeliac disease locus in the Dutch population. Independent confirmation of involvement of the 19p13.1 locus in coeliac disease was provided by the presence of association within this region in a different data set. Furthermore, suggestive linkage to the 6q21-22 region was detected. The 19p13.1 and 6q21-22 loci both confer a considerable risk for coeliac disease development, with a λ_s of 2.6 and 2.3 respectively. None these loci are age-dependent. The next step is to identify these important coeliac disease genes. Recently, positional cloning of genes involved in complex diseases has proved to be successful. In 2000, the calpain-10 gene was identified as the type 2 diabetes mellitus susceptibility gene in the NIDDM1 region on chromosome 2.[30] Furthermore, variants in the CARD15 (previously named NOD2) gene on chromosome 16 were found to be associated with Crohn's disease[31] and the ADAM33 gene on chromosome 20 was recently identified as an asthma susceptibility gene.[32] The 6q21-22 and 19p13.1 regions are still too large for candidate gene studies, and further

fine-mapping of the region is underway. Identification and characterization of these coeliac disease genes will lead to important new insights in the pathogenesis of coeliac disease and may open new opportunities for diagnosis. First of all, when the coeliac disease susceptibility alleles of these genes have been determined, more accurate risk estimations can be offered to relatives of Dutch coeliac disease patients, based on the genotype at HLA-DQA1 and DQB1, 19p13.1 and 6q21-22. This may also be helpful in diagnosing patients with a borderline pathology or patients who were already on a gluten-free diet at the time of the first biopsy. Secondly, when the gene is known, functional studies can be performed to determine the role of this protein. Unravelling the biological pathway in which this protein acts will lead to understanding of the disease process and possibly to a target for therapeutic intervention.

Acknowledgements

This study is dedicated to the memory of Lodewijk Sandkuijl (1953-2002), who died shortly after its completion. He was an inspiration to us, and a world expert on biostatistics.

We are grateful to the Dutch Coeliac disease Foundation for helping us contact the coeliac disease families and to all the families who participated in our study. The authors wish to thank Arjenne Kors, Anath Oren and Winny van der Hoeven for collecting the blood samples, Eric Strengman and Ruben van 't Slot for isolating the DNA and checking the blind control samples, Lude Franke for blasting the primer sequences, Alienke Monsuur for typing part of the chromosome 6 markers and Bobby Koeleman for statistical advice. We also thank Greetje van Berlo-Jansen for sending all the biopsy reports, Salvador Peña for his valuable suggestions regarding the manuscript and Jackie Senior for improving the manuscript.

References

1. Farrell RJ, Kelly CP. Diagnosis of coeliac sprue. *Am J Gastroenterol* 2001; **96:** 3237-46.
2. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('coeliac sprue'). *Gastroenterology* 1992; **102:** 330-54.
3. Sollid LM. Molecular basis of coeliac disease. *Annu Rev Immunol* 2000; **18:** 53-81.
4. Greco L, Romino R, Coto I*, et al.* The first large population based twin study of coeliac disease. *Gut* 2002; **50:** 624-8.

5. Sollid LM, Thorsby E. HLA susceptibility genes in coeliac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 1993; **105:** 910-22.
6. Schuppan D. Current concepts of coeliac disease pathogenesis. *Gastroenterology* 2000; **119:** 234-42.
7. Woolley N, Holopainen P, Ollikainen V*, et al.* A new locus for coeliac disease mapped to chromosome 15 in a population isolate. *Hum Genet* 2002; **111:** 40-5.
8. Rostami K, Kerckhaert J, Tiemessen R, von Blomberg BM, Meijer JW, Mulder CJ. Sensitivity of antiendomysium and antigliadin antibodies in untreated coeliac disease: disappointing in clinical practice. *Am J Gastroenterol* 1999; **94:** 888-94.
9. Carrington M, Miller T, White M*, et al.* Typing of HLA-DQA1 and DQB1 using DNA single-strand conformation polymorphism. *Hum Immunol* 1992; **33:** 208-12.
10. Greco L, Babron MC, Corazza GR*, et al.* Existence of a genetic risk factor on chromosome 5q in Italian coeliac disease families. *Ann Hum Genet* 2001; **65:** 35-41.
11. Greco L, Corazza G, Babron MC*, et al.* Genome search in coeliac disease. *Am J Hum Genet* 1998; **62:** 669-75.
12. Liu J, Juo SH, Holopainen P*, et al.* Genome-wide linkage analysis of coeliac disease in Finnish families. *Am J Hum Genet* 2002; **70:** 51-9.
13. Naluai AT, Nilsson S, Gudjonsdottir AH*, et al.* Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for coeliac disease on chromosomes 5 and 11. *Eur J Hum Genet* 2001; **9:** 938-44.
14. King AL, Moodie SJ, Fraser JS*, et al.* CTLA-4/CD28 gene region is associated with genetic susceptibility to coeliac disease in UK families. *J Med Genet* 2002; **39:** 51-4.
15. Holopainen P, Mustalahti K, Uimari P, Collin P, Maki M, Partanen J. Candidate gene regions and genetic heterogeneity in gluten sensitivity. *Gut* 2001; **48:** 696-701.
16. Zhong F, McCombs CC, Olson JM*, et al.* An autosomal screen for genes that predispose to coeliac disease in the western counties of Ireland. *Nat Genet* 1996; **14:** 329-33.
17. Risch N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 1990; **46:** 229-41.
18. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; **11:** 241-7.
19. Kruglyak L, Lander ES. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 1995; **57:** 439-54.
20. Holmans P. Asymptotic properties of affected-sib-pair linkage analysis. *Am J Hum Genet* 1993; **52:** 362-74.
21. Risch N. Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet* 1987; **40:** 1-14.
22. Bevan S, Popat S, Braegger CP*, et al.* Contribution of the MHC region to the familial risk of coeliac disease. *J Med Genet* 1999; **36:** 687-90.
23. Houlston RS, Ford D. Genetics of coeliac disease. *Qjm* 1996; **89:** 737-43.
24. Popat S, Bevan S, Braegger CP*, et al.* Genome screening of coeliac disease. *J Med Genet* 2002; **39:** 328-31.
25. King AL, Fraser JS, Moodie SJ*, et al.* Coeliac disease: follow-up linkage study provides further support for existence of a susceptibility locus on chromosome 11p11. *Ann Hum Genet* 2001; **65:** 377-86.
26. Rioux JD, Silverberg MS, Daly MJ*, et al.* Genome-wide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* 2000; **66:** 1863-70.
27. Pociot F, McDermott MF. Genetics of type 1 diabetes mellitus. *Genes Immun* 2002; **3:** 235-49.

28. Gaffney PM, Ortmann WA, Selby SA*, et al.* Genome screening in human systemic lupus erythematosus: results from a second Minnesota cohort and combined analyses of 187 sib-pair families. *Am J Hum Genet* 2000; **66:** 547-56.
29. Veal CD, Clough RL, Barber RC*, et al.* Identification of a novel psoriasis susceptibility locus at 1p and evidence of epistasis between PSORS1 and candidate loci. *J Med Genet* 2001; **38:** 7-13.
30. Horikawa Y, Oda N, Cox NJ*, et al.* Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 2000; **26:** 163-75.
31. Hugot JP, Chamaillard M, Zouali H*, et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **411:** 599-603.
32. Van Eerdewegh P, Little RD, Dupuis J*, et al.* Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* 2002; **418:** 426-30.
33. King AL, Yiannakou JY, Brett PM*, et al.* A genome-wide family-based linkage study of coeliac disease. *Ann Hum Genet* 2000; **64:** 479-90.
34. Neuhausen SL, Feolo M, Camp NJ, Farnham J, Book L, Zone JJ. Genome-wide linkage analysis for coeliac disease in North American families. *Am J Med Genet* 2002; **111:** 1-9.

Chapter 3

A genome-wide screen in a large family

A genome-wide screen in a four-generation Dutch family with coeliac disease: evidence for linkage to chromosomes 6 and 9

M.J. van Belzen, M.M. Vrolijk, J.W.R. Meijer, J.B.A. Crusius, P.L. Pearson, L.A. Sandkuijl, R.H.J. Houwen and C. Wijmenga

American Journal of Gastroenterology; in press.

Summary

Objectives: Coeliac disease is caused by the interaction of multiple genes and environmental factors. Inheritance of the disease shows a complex pattern with a 10% sibling recurrence risk. The HLA-region is a major genetic risk locus in coeliac disease, but genes outside this region are expected to contribute to the disease risk as well. The aim of this study was to identify the loci causing coeliac disease in one large Dutch family with apparent dominant transmission of the disease. *Methods:* The family comprised 17 patients in four generations, with possible transmission of the disease by both grandparents. Microsatellite markers evenly spread over all chromosomes were genotyped and linkage analysis was performed using both dominant and recessive disease models and a model-free analysis. *Results:* Disease susceptibility in the family was linked to the HLA-region (lod score of 2.33) and all patients were HLA-DQ2. A dominantly inherited non-HLA locus with a maximum lod score of 2.61 was detected at 9p21-13, which was shared by 16 patients. Model-free analysis identified another possible non-HLA locus, at 6q25.3, which was shared by 14 patients ($p = 0.01$). Neither of these regions were detected in a genome-wide screen in Dutch affected sibpairs, but the 9p21 locus has been implicated in Scandinavian families. *Conclusions:* Two potential non-HLA loci for coeliac disease were identified in this large Dutch family. Our results provide replication of the Scandinavian 9p21 locus, and suggest that this locus plays a role in coeliac disease patients from different Caucasian populations.

Introduction

Coeliac disease is an autoimmune disorder with a multifactorial aetiology. Ingestion of gluten, present in wheat, barley and rye, leads to a range of aberrations in the small intestinal mucosa of these patients. The range of abnormalities can be classified according the modified Marsh classification. The Marsh type III lesion is the most severe lesion, with the presence of intra-epithelial lymphocytosis with damage of the mucosal epithelial cells, crypt hyperplasia and partial (IIIa), subtotal (IIIb) or total (IIIc) villous atrophy.[1,2] A wide spectrum of clinical symptoms can be associated with this lesion, including diarrhoea, abdominal pain and bloating. Symptoms due to malabsorption of nutrients, like fatigue, weight loss, anaemia and osteopenia, are also frequently present. Most coeliac disease patients that adhere strictly to the gluten-free diet (GFD) show improvement or

complete disappearance of the clinical symptoms and recovery of the small intestinal mucosa.[3] Recently, it has been recognized that less severe lesions, like lymphocytosis without or with crypt hyperplasia (Marsh types I and II, respectively), can also be associated with the same clinical spectrum.[2-4]

Coeliac disease is a strongly inheritable disorder, with a relative risk of approximately 10% for siblings of a patient.[1] One important genetic factor is the HLA-region, with the majority of patients expressing HLA-DQ2, and almost all of the remaining patients expressing HLA-DQ8.[5] However, the genetic contribution of the HLA-region to coeliac disease has been estimated at only ~40%.[6,7] Therefore, non-HLA genes must also contribute to the disease.

Identification of susceptibility genes for multifactorial disorders is hampered by multigenic aetiology and genetic heterogeneity.[8] It is expected that a causative variant in a susceptibility gene for these diseases will be common in the population, and most carriers of this variant will not develop the disease since they do not carry all the necessary disease susceptibility genes to pass the disease threshold. Furthermore, different genes can cause identical phenotypes in different families. To overcome these difficulties, large numbers of cases and controls are needed in association studies, and large numbers of families with multiple patients are needed for linkage analysis. Within a single family the disease is expected to be genetically homogeneous, as the same genes are likely to cause the disease in all patients from that family. Therefore, a large family with many patients in different generations may provide a unique opportunity to identify genes that cause these multifactorial diseases.

We present here a family with 17 coeliac disease patients in four generations. Ten out of thirteen (76%) siblings from the second generation have coeliac disease, which is much more than would be expected from the average sibling relative risk. We therefore hypothesized that a single non-HLA gene, with high penetrance, causes coeliac disease in this family and a genome-wide screen was performed to localize this gene locus.

Subjects and methods

The study family

This study was approved by the Medical Ethics Committee of the University Medical Centre Utrecht and written informed consent was obtained from all the participants. The

family is of Dutch origin and lives in the north of the Netherlands. At the time of ascertainment, only five patients were diagnosed with coeliac disease: individuals (IDs) 05, 09, 14, 19 and 20 (Figure 1). Serological screening was offered to 36 additional family members and they were screened for antigliadin IgA (AGA), antigliadin IgG (AGG) and anti-endomysium (Ema) antibodies. Anti-tissue transglutaminase (tTG) antibodies could only recently be determined, and this test was performed in only three individuals. Twelve additional coeliac disease patients were diagnosed by a duodenal biopsy. All biopsy specimens, including those from the initial patients, were retrieved for re-evaluation and Marsh classification by one experienced pathologist (JWRM). The histological lesions were classified according to Marsh.[1] The Marsh III category was subdivided into three groups; i.e. partial (Marsh IIIa), subtotal (Marsh IIIb) and total (Marsh IIIc) villous atrophy.[9] All patients were genotyped at the HLA-DQA1 and DQB1 loci as described before.[10] An overview of symptoms, serology, and Marsh classification before and on a GFD, together with the HLA-DQ type, is given in Table 1. Serology after GFD was not available.

Linkage analysis was performed using an "affecteds-only" strategy, so only twenty-three individuals were included in this study: the 17 coeliac disease patients, five spouses and the grandfather (Figure 1 and Table 1). Twelve of the patients had a Marsh III lesion and all of them responded well to the GFD. Three individuals, IDs 02, 19 and 24, showed a Marsh II lesion. The grandmother (ID 02) does not adhere to a GFD, but the presence of the coeliac disease-specific Ema antibodies, the Marsh II lesion and her clinical symptoms strongly suggest coeliac disease. Both IDs 19 and 24 showed significant clinical improvement after only two weeks on a GFD and they experience strong reactions to accidental gluten intake. The Marsh II lesion from ID 19 was normalized to Marsh 0 after two years on a GFD. Biopsy specimens from the proband (ID20) and ID14 reportedly showed villous atrophy, but could not be retrieved for re-evaluation because the diagnoses were made in 1978 and 1980, respectively. ID 20 was hospitalised at the age of one for severe weight loss and diarrhoea, but recovered completely on a GFD; ID 14 also experienced serious complaints which all disappeared on a GFD. The grandfather (ID 01) was positive for AGA and AGG, but negative for the more coeliac disease specific Ema and tTG antibodies. He suffers from dermatitis herpetiformis and fatigue, but he refused a duodenal biopsy.

Table 1. Patient characteristics.

ID[a]	Year of Birth	Symptoms before GFD[b]	Serology[c]	Marsh before GFD[d]	DQ type	Symptoms after GFD	Marsh after GFD[d]
01	1920	fatigue, DH, cannot eat bread	AGA, AGG	ND	DQ2	no GFD	
02	1923	fatigue, anaemia	AGA, Ema	II	DQ2	no GFD	
05	1947	fatigue, anaemia, weight loss, abdominal pain/bloating	ND	IIIb	DQ2	disappeared	IIIa
07	1949	fatigue, diarrhoea, irritable	AGA, Ema	IIIc	DQ2	disappeared	ND
08	1950	fatigue, abdominal pain	AGA, Ema	IIIa	DQ2	disappeared	ND
09	1952	diarrhoea, abdominal pain	ND	IIIb	DQ2	disappeared	ND
10	1953	fatigue	Ema	IIIa	DQ2	improved	ND
11	1954	fatigue, diarrhoea	AGA, AGG, Ema	IIIb	DQ2	improved	ND
12	1955	fatigue, abdominal pain/bloating	AGA, Ema	IIIa	DQ2	disappeared	II
13	1957	fatigue, abdominal bloating, irritable	AGA	IIIa	DQ2	disappeared	ND
14	1960	fatigue, diarrhoea, anaemia, weight loss, DH	ND	NA	DQ2	disappeared	ND
16	1962	fatigue, diarrhoea	Ema	IIIa	DQ2	improved	II
19	1971	fatigue, bleeding gums	ND	II	DQ2	disappeared	0
20	1976	fatigue, diarrhoea, weight loss, irritable, tooth enamel defects	ND	NA	DQ2	disappeared	0
24	1978	fatigue	Ema	II	DQ2	disappeared	ND
32	1989	fatigue	Ema	IIIb	DQ2	disappeared	ND
41	1998	fatigue, irritable	AGA, Ema, tTG	IIIb	DQ2	disappeared	ND
55	1978	none	Ema	IIIb	DQ2	no GFD	

[a] ID refers to the identification numbers used in Figure 1.
[b] GFD = gluten-free diet; DH = Dermatitis Herpetiformis.
[c] AGA = antigliadin IgA; AGG = antigliadin IgG; Ema = anti-endomysium IgA; ND = not done and tTG = anti-tissue transglutaminase IgA. Note: tTG testing was only performed in ID 01 and 41.
[d] NA = not available; ND = not done.

Genotyping of microsatellite markers

A genome-wide screen was performed in all twenty-three individuals (Figure 1). A total of 321 microsatellite markers were genotyped, evenly spread over the genome with an average distance of 15 cM. A marker spacing <5 cM was obtained for regions that had been implicated in our genome-wide screen in Dutch affected sibpairs (6q21-22 and 19p13.1)[11] and regions showing linkage in other studies (*CELIAC1* in the HLA-region, *CELIAC2* at 5q31-33 and *CELIAC3* in the *CTLA4/CD28* region at 2q33-34). Regions with lod scores >1.0 were selected for fine-mapping, and 50 additional markers were

genotyped in these regions. Our marker set consisted mainly of screening set 6 from the Marshfield Centre for Medical Genetics, complemented with markers selected from the Marshfield and Ensembl genetic maps. DNA from the family members was amplified by PCR, together with three CEPH reference samples and a negative control. The reaction volume of 10 µl contained 25 ng of DNA, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 50 ng fluorescence-labelled primer and 0.4 U AmpiTaq Gold (PE Applied Biosystems, Foster City, CA). The PCR products were pooled and separated on a 3700 DNA sequencer (PE Applied Biosystems) and analysed by Genescan 3.5 and Genotyper 2.0 software (PE Applied Biosystems). All genotypes were checked independently by two researchers. The identity of the markers was verified by comparing genotypes of the CEPH reference samples to the CEPH genotype database. A Mendelian inheritance check was performed and markers with Mendelian errors were excluded from the linkage analysis.

Linkage analysis

A parametric linkage analysis was performed with the MLINK program of the LINKAGE package[12] using an "affecteds-only" approach. This approach is preferred in complex diseases, as it is possible that healthy individuals also carry some of the disease susceptibility loci. All 23 individuals who were genotyped were included in the linkage analysis and the disease status of the 17 patients, either with a Marsh II lesion, a Marsh III lesion or with unavailable biopsy specimens, was affected. The five spouses were not tested for coeliac disease and were labelled unknown. The disease status of the grandfather (ID 01) was also unknown, as coeliac disease was not histologically proven in this individual.

A two-point linkage analysis was performed for all markers using a dominant inheritance model as well as a recessive model (parametric linkage analysis). A multipoint analysis was performed in regions with a lod score >1.0 using the FASTLINK program. For these analyses, the model that produced linkage in the two-point analysis was used. A maximum of three successive markers was included in the parametric multipoint analysis because of calculation time considerations. Power calculations were performed with the SLINK and MSIM programs. For all analyses, the penetrance for the normal genotype was equal to the population frequency of coeliac disease (0.005) and the penetrance of the disease genotype was 0.8. Disease allele frequency was 0.001, based on the hypothesis

that a rare variant of a single susceptibility gene is causing the exceptionally large number of patients in this family. Allele frequencies were set equal for all alleles, since all the parental genotypes were available and therefore the allele frequencies in the datafile were not used in the analysis. The order and location of the markers on the chromosomes was based on the Marshfield genetic map.

Although the disease transmission in this family suggests autosomal dominant inheritance, other models cannot be excluded. To check for the possibility that linkage was missed because of applying a wrong model, a model-free (non-parametric) analysis was also performed using the Genehunter program.[13] When calculating the non-parametric lod score (NPL) statistic, a disease model is not used but the sharing of marker alleles by all possible pairs of affected individuals is determined and compared to the expected values based on the familial relationship. The study family is too large for the Genehunter program, and was therefore divided in two smaller families (family A: IDs 01, 02, 05, 07, 08, 09, 18, 19, 20, 21, 24, 40 and 41; family B: IDs 01, 02, 10, 11, 12, 13, 14, 16, 31, 32, 54 and 55). For this purpose, the affection status of both grandparents (ID 01 and 02) was set as unknown. NPL values obtained for both family branches were totalled.

Results

To determine the power of the study family, simulation studies were performed using the same parameters as selected for the linkage analysis. The maximum attainable lod score was 4.17 assuming a dominant model of transmission and 3.65 assuming a recessive model. These simulation lod scores were obtained with 100% informativity of the marker and no recombination between the marker and the disease locus. The probability of obtaining lod scores of at least 1.0, 2.0 or 3.0 were 53%, 25% and 8% respectively. Because of the low probability of obtaining high lod scores, all regions with a lod score >1.0 were selected for fine-mapping.

Two-point lod scores were calculated for all 321 microsatellite markers and seven regions with a lod score >1 were identified, on chromosomes 3, 6, 9, 10, 15, 16 and 19. Fifty additional markers were typed in these regions, and the highest lod scores obtained in each region are shown in Table 2, together with the p-value corresponding to the NPL statistic from the two-point model-free analysis.

Table 2. Regions with two-point lod scores >1.0.

Location	Position[a]		Marker[b]	Lod score[c]	NPL p-value[d]	Information content
	cM	Mb				
3q12.3	119.1	101.6	D3S2459	1.13	0.02	0.91
6p21.3	–	31.3	MIB	2.33	0.01	0.95
	45.0	31.7	D6S273	1.79	0.02	0.87
9p21-13	55.3	31.1	D9S43	1.55	0.03	0.82
	58.3	32.1	D9S1118	1.64	0.02	1.0
	59.3	34.0	D9S1817	2.61	0.0005	1.0
	–	35.3	D9S163	2.61	0.0003	0.66
	59.9	36.1	D9S1804*	0.78	0.19	0.52
	60.6	36.9	D9S50	2.61	0.0003	0.66
	61.4	37.4	D9S1874*	0.80	0.16	0.52
	–	38.2	D9S200	1.43	0.02	0.95
10q26.2	156.2	–	D10S1223	1.33	0.23	1.0
	160.0	129.1	D10S1676	1.33	0.08	0.82
15q22.3	62.4	59.7	D15S153*	1.36	0.12	0.61
16q23.2	108.3	82.4	D16S3098	1.17	0.004	0.66
	111.1	83.9	D16S422	1.17	0.23	0.66
19p13.1	45.5	17.6	D19S899	1.17	0.42	1.0
	47.3	18.3	D19S915	1.17	0.42	1.0

[a] The genetic locations are based on the Marshfield genetic map and the physical locations are based on the Ensembl map (March 2003 release). Markers without a position were not present in that map.
[b] Markers marked by an asterisk were not informative in ID 02.
[c] Lod scores are based on the dominant transmission model, except for the markers at 10q26.2 which were obtained with the recessive model.
[d] P-value corresponding to the NPL statistic of the non-parametric linkage analysis.

The HLA region on chromosome 6 was linked to coeliac disease in this family, with a lod score of 2.33 at marker MIB. Multipoint linkage analysis resulted in the same lod score. The most promising non-HLA region is located on chromosome 9, which reached a maximum lod score of 2.61 at markers D9S50, D9S1817 and D9S163. D9S1817 was completely informative in all meioses, so the maximum multipoint lod score in this region was also 2.61. Model-free analysis also demonstrated increased allele sharing among the affected individuals, with an NPL score of 7.81 ($p = 0.0005$) at D9S1817. Haplotypes of eight markers from this region are depicted in Figure 1. Sixteen out of 17 patients share a five-marker haplotype ranging from D9S1817 to D9S1874. Only one patient, ID 11, inherited the other haplotype from the grandmother. The maximum size of this candidate region is 6.1 Mb and is defined by markers D9S1118 and D9S200. Multipoint linkage analysis of the 3q12.3, 10q26.2, 16q23.2 and 19p13.1 regions resulted in lod scores identical to the single point lod scores. A lower multipoint lod score was obtained for 15q22.3 (0.67).

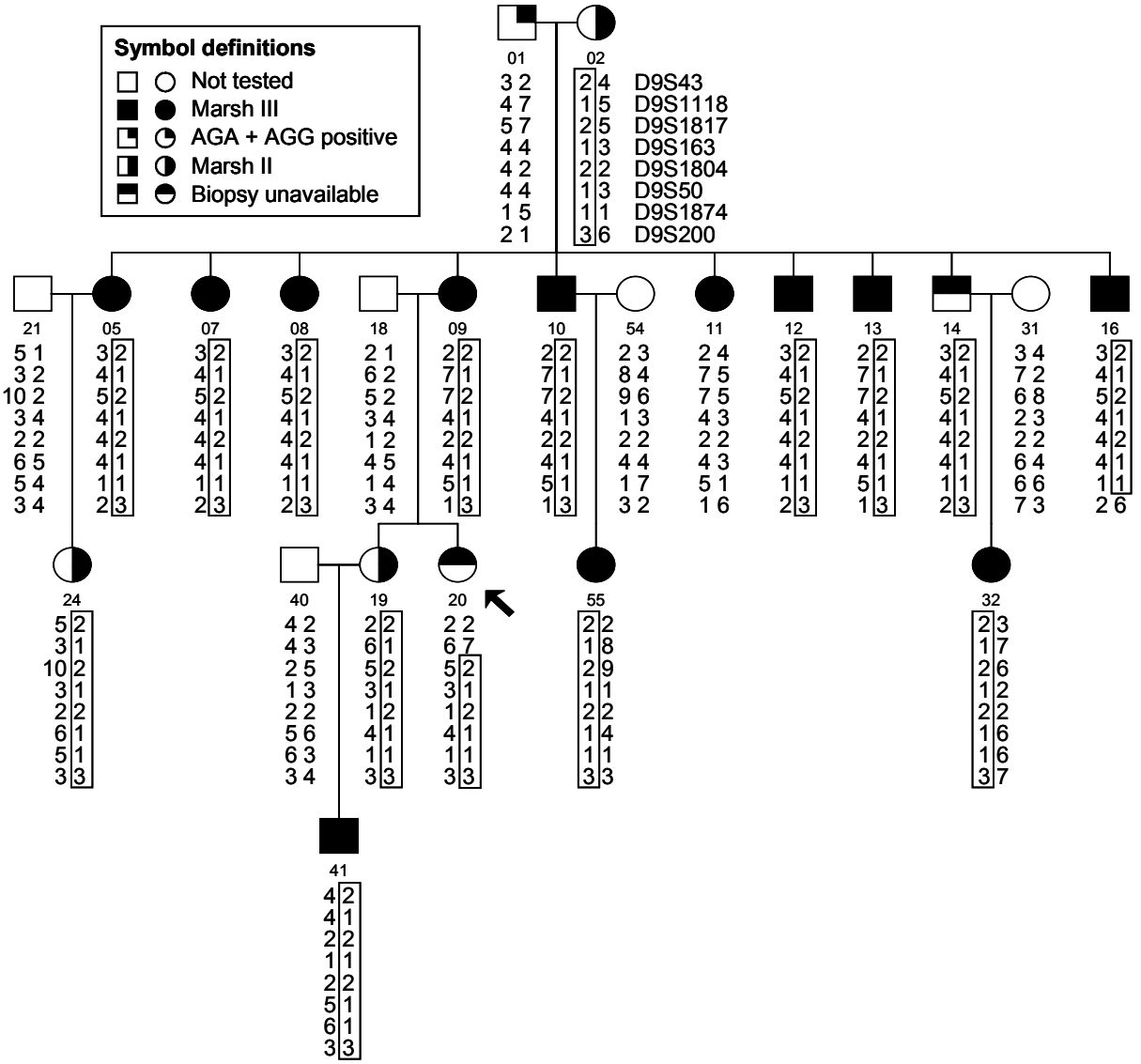


Figure 1. Pedigree of the study family and chromosome 9p21-13 haplotypes.

Only affected individuals are depicted. The disease status of individuals with open symbols and the grandfather (ID 01) was unknown, all others were considered to be affected. The region on chromosome 9 that is shared by all affected individuals, except ID 11, is boxed. The arrow indicates the proband.

Non-parametric linkage analysis identified three additional regions with nominal p-values <0.05 on chromosomes 6, 10 and 11. The NPL value at 6q25.3 was 4.58 (p = 0.01) and peaked at marker D6S969. This locus was inherited from the grandfather (ID 01) and is shared by 14 of 16 affected offspring. Only IDs 13 and 16 inherited the other haplotype. Marker D10S1227 at 10q21.1 produced an NPL value of 4.22 (p = 0.02), this

locus was inherited from the grandmother (ID 02) and was transmitted to 11 of 16 offspring. On chromosome 11 at p15.4, an NPL value of 5.49 ($p = 0.002$) was obtained at markers D11S2362 and D11S1760. This locus is shared by 10 offspring and was inherited from the grandfather. It is present in all individuals in branch A of the family. However, only 2 individuals in branch B, IDs 11 and 13, also carry this haplotype. The high NPL value of this locus was contributed only by branch A, resulting in a biased total NPL value for the entire family.

Discussion

A Dutch family with 17 coeliac disease patients in four generations is presented in this paper. Although the family was already known with coeliac disease since 1978, only five patients were diagnosed by 2001. Intensive screening of the other family members resulted in the identification of 12 additional patients. Before diagnosis, most of these 12 patients did not report complaints. But after only a few weeks of adhering to a GFD, most of them reported a dramatic decrease of fatigue and disappearance of mild abdominal pain and diarrhoea. All these individuals stated that their quality of life had increased significantly and they are willing to adhere to the GFD in the future. This emphasizes once again the importance of screening for coeliac disease in family members of patients, even when no obvious complaints are reported.

The family presented here was originally collected for our affected sibpair study.[11] However, the family was excluded because of the exceptionally large number of affected individuals. This family alone would have provided 46 sibpairs, compared to 84 from the other 67 families and this would have seriously biased our results. One explanation for the high percentage of affected offspring in the second generation could be consanguinity between the grandparents, resulting in transmission of the same disease locus by both grandparents. The grandparents (IDs 01 and 02) originated from two small neighbouring villages in the north of the Netherlands, but genealogical investigation did not reveal consanguinity within the last six generations.

The apparent dominant disease transmission through four generations suggested a Mendelian cause of the disease in this family. Surprisingly however, bilineal transmission of two possible non-HLA disease loci, at 9p21-13 and 6q25.3 respectively, was detected, and neither of these loci was shared by all patients. Sixteen affected individuals share a

haplotype of 2.1 cM at 9p21-13, which was transmitted by the affected grandmother (ID 02). Only one patient, ID 11, inherited the other maternal haplotype. There are no reasons to doubt her diagnosis, so her disease is probably caused by other susceptibility genes. In addition, a locus at 6q25.3 was inherited from the grandfather (ID 01) and shared by 14 of 16 affected offspring. Transmission of a disease locus by the grandfather is not at all unlikely, as he presents with several characteristics compatible with coeliac disease. The 6q25.3 locus was missed in the parametric analysis, because of the model that was used. This locus is present in ID 11, possibly compensating for the absence of the 9p21-13 locus. Several other loci with parametric lod scores >1.0 and non-parametric p-values <0.05 were also present, and some of these may add to the disease risk in some patients.

Recently we have completed a genome-wide screen in affected Dutch MIII sibpairs.[11] We identified a major coeliac disease locus on chromosome region 19p13.1, and a second locus on chromosome region 6q21-22, distinct from HLA. No evidence for linkage to 9p21-13 or 6q25.3 was obtained in our affected sibpair study. So the major loci in this large family are not important loci in Dutch coeliac disease patients in general. On the other hand, the two major loci identified in the affected sibpair study have no significant effect in this family. However, further evidence for linkage of coeliac disease to 9p21 was provided by two previous studies in Swedish/Norwegian ($p = 0.038$)[14] and Finnish affected sibpairs (lod = 1.11)[15], indicating that this locus may be a true, but not a major, risk factor for coeliac disease in general. Linkage of coeliac disease to 6q25.3 was not present in other populations, but linkage to this region has been observed several times in type 1 diabetes mellitus.[16] Type 1 diabetes mellitus is associated with coeliac disease,[17] and this locus may harbour a shared susceptibility gene for these disorders.

As expected, linkage to the HLA-region at 6p21.3 was also present, with a maximum lod score of 2.33 at marker MIB. Although all affected individuals in the study family are DQ2-positive and this marker is informative in all but one transmission, the maximum attainable lod score was not reached. Both grandparents are heterozygous DQ2 carriers, and ID 14 is homozygous DQ2. He has transmitted the grandfather's DQ2 haplotype to his affected daughter (ID 32), thereby destroying the linkage. This situation illustrates the problems frequently encountered in linkage analysis in complex disorders. When risk alleles are common in the population, they can also be present in

healthy spouses. In this study family, DQ2 is also married-in by ID 18 (DQ2 heterozygote) and ID 40 (DQ2/DQ8).

Even within this one family, the disease appears to be genetically heterogeneous and of multigenic origin. Surprisingly, the major loci identified in our affected sibpair study do not play a significant role in this family. Nevertheless, this large family may be important for understanding of the pathogenesis of coeliac disease. The loci involved in this family may not be of major importance for coeliac disease in general, but they may play a role in a small proportion of coeliac disease patients. These loci may be hard to detect by an affected sibpair approach because of the small relative risk of such loci.

Acknowledgements

The authors are very grateful to all members of the study family, who participated with great enthusiasm and willingness in this study. We wish to thank Winny van der Hoeven for her help in collecting blood samples and clinical information from the family members. Jackie Senior is acknowledged for critically reading the manuscript and Eric Hennekam for the genealogical investigation. The study was financially supported by a grant from the Dutch Digestive Diseases Foundation (WS 97-44).

This paper is dedicated to the memory of Lodewijk Sandkuijl. He participated in this project from the beginning and his contribution to the design of the study and analysis of the data was of great importance to us. He died unexpectedly, shortly before the genome-wide screen was completed.

References

1. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 1992; **102:** 330-54.
2. Mulder CJ. When is a coeliac a coeliac? Report of a working group of the United European Gastroenterology Week in Amsterdam, 2001. *Eur J Gastroenterol Hepatol* 2001; **13:** 1123-8.
3. Farrell RJ, Kelly CP. Diagnosis of celiac sprue. *Am J Gastroenterol* 2001; **96:** 3237-46.
4. Wahab PJ, Crusius JB, Meijer JW, Mulder CJ. Gluten challenge in borderline gluten-sensitive enteropathy. *Am J Gastroenterol* 2001; **96:** 1464-9.
5. Sollid LM. Molecular basis of celiac disease. *Annu Rev Immunol* 2000; **18:** 53-81.
6. Bevan S, Popat S, Braegger CP*, et al.* Contribution of the MHC region to the familial risk of coeliac disease. *J Med Genet* 1999; **36:** 687-90.
7. Petronzelli F, Bonamico M, Ferrante P*, et al.* Genetic contribution of the HLA region to the familial clustering of coeliac disease. *Ann Hum Genet* 1997; **61:** 307-17.

8. Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000; **405:** 847-56.
9. Rostami K, Kerckhaert J, Tiemessen R, von Blomberg BM, Meijer JW, Mulder CJ. Sensitivity of antiendomysium and antigliadin antibodies in untreated celiac disease: disappointing in clinical practice. *Am J Gastroenterol* 1999; **94:** 888-94.
10. Wahab PJ, Crusius JB, Meijer JW, Uil JJ, Mulder CJ. Cyclosporin in the treatment of adults with refractory coeliac disease-- an open pilot study. *Aliment Pharmacol Ther* 2000; **14:** 767-74.
11. van Belzen MJ, Meijer JW, Sandkuijl LA*, et al.* A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* 2003; **125:** 1032-41.
12. Lathrop GM, Lalouel JM, Julier C, Ott J. Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci U S A* 1984; **81:** 3443-6.
13. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58:** 1347-63.
14. Naluai AT, Nilsson S, Gudjonsdottir AH*, et al.* 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* 2001; **9:** 938-44.
15. Liu J, Juo SH, Holopainen P*, et al.* Genome-wide linkage analysis of celiac disease in Finnish families. *Am J Hum Genet* 2002; **70:** 51-9.
16. Pociot F, McDermott MF. Genetics of type 1 diabetes mellitus. *Genes Immun* 2002; **3:** 235-49.
17. Holmes GK. Screening for coeliac disease in type 1 diabetes. *Arch Dis Child* 2002; **87:** 495-8.

Part II

Association analysis

Chapter 4

Fine-mapping of the chromosome 19 region

Fine-mapping of the coeliac disease region on chromosome 19

M.J. van Belzen, A. Zhernakova, A.F.J. Bardoel, S.C. Bakker, P.L. Pearson, R.H.J. Houwen and C. Wijmenga

Preliminary results.

Summary

Coeliac disease is a common autoimmune disorder, which is strongly associated to the HLA region. However, the contribution of the HLA region has been estimated to be no more than 50%, so non-HLA genes must also be involved in the aetiology of coeliac disease. We have recently performed a genome-wide screen to localise these genes, and we identified significant linkage to 19p13.1. This region was subjected to systematic fine-mapping in a case-control cohort, using a DNA pooling strategy, to identify the coeliac disease susceptibility gene. Firstly, the region was saturated with microsatellite markers and strongest association was detected at marker D19S899 (overall P-value = 0.0013), which is the marker at which the multipoint lod score peaked. The 450 kb region of interest around this marker was covered with single nucleotide polymorphisms (SNPs), which were typed in the patient and control pools. Eleven SNPs showed significant association, four of which were subsequent SNPs in the 3' region of the gene encoding myosin IXB (*MYO9B*). Single typing of 15 SNPs in the *MYO9B* gene region showed association of six of them (P <0.01). Five SNPs in the *MYO9B* gene combine into a haplotype, which is present in 38.8% of patients and in 30.9% of controls (P = 0.02). However, the associated region may actually cover eight genes and further research should be directed towards the extent of the observed association before mutational analysis of positional candidate genes can be commenced.

Introduction

The development of coeliac disease is clearly influenced by genetic factors and the importance of HLA-DQ2 and DQ8 has been known for a long time. Furthermore, it has been well established that other, non-HLA genes must contribute to the disease as well. Ten genome-wide screens have been completed so far to localise these genes (see also Chapters 2 and 3). However, no non-HLA genes contributing to coeliac disease have been identified to date, although there is growing evidence for a role of the *CTLA4* gene.[1]

We have recently completed a genome-wide screen in 101 Dutch sibpairs affected with coeliac disease and identified significant linkage to chromosome region 19p13.1 (MMLS 4.31; P = 6.2 x 10^{-6}, see also Chapter 2).[2] This locus makes a considerable contribution to coeliac disease, with an estimated sibling relative risk of 2.6. The

identical-by-descent (IBD) allele sharing at the peak of the multipoint lod score was Z_0 = 0.095, Z_1 = 0.5, Z_2 = 0.405 ($Z_{0,1,2}$ represents the proportion of affected sibpairs sharing 0, 1 or 2 alleles IBD, respectively (data not shown)). This IBD distribution suggests that the 19p13.1 locus acts in a dominant mode with approximately 60% of the families linked to this locus, thus providing us with an excellent opportunity to identify this important susceptibility gene for coeliac disease.

The 19p13.1 candidate region, defined by the maximum lod score –1 support interval, spans some 3 Mb (based on the linkage analysis in all 82 families). However, this region is very gene-rich and contains 92 genes and ESTs. To identify the coeliac disease susceptibility gene, we first performed fine-mapping of the region using a DNA-pooling strategy in order to narrow-down the number of candidate genes. SNPs within the associated region were subsequently tested for association with coeliac disease.

Subjects and Methods

Subjects

DNA, isolated from whole blood or buccal cells, was available from 216 independent Dutch coeliac disease patients. The initial biopsy specimens of the patients were retrieved and all showed a Marsh III lesion upon re-evaluation by one experienced pathologist (Dr. Meijer, Rijnstate Hospital, Arnhem). The cohort consisted of children and adults, with 65% females and a mean age of 39 years. There were also 216 Dutch controls available, matched for age and sex to the patients. In addition, 122 parent-case trios, with one child with a Marsh III lesion and both parents, were available for TDT. The study was approved by the Medical Ethics Committee of the University Medical Centre Utrecht and written informed consent was obtained from the coeliac disease patients and the control individuals.

Preparation of DNA pools

DNA pools were constructed by pooling equal amounts of genomic DNA isolated from peripheral blood lymphocytes from patients and controls. DNA isolated from buccal cells was excluded because of its lower quality. Therefore, 199 coeliac disease patients were included and divided into two patient pools. Two control pools were created, comprising 199 matched control individuals. The DNA pools were constructed

according to standard protocols.[3] In short, the DNA concentration of the individual samples was measured with PicoGreen and 200 ng of DNA of each individual was added to the appropriate pool. Pools were purified by phenol extraction and dissolved in water to a final concentration of 15 ng/μl.

Genotyping of microsatellite markers

All known microsatellite markers within the candidate region (based on the Ensembl map of November 2002) were genotyped in the patient and control pools, giving an average distance of 250 kb (Table 1). Several markers adjacent to this region were also typed to ensure that the entire region that could harbour the gene was covered, as the 3 Mb region is merely a support interval, based on the results in all 82 families (which is smaller than the support interval estimated from the results in the initial 67 families). Five of the 23 markers included in this study were also used for the linkage analysis in affected sibpairs.[2] The markers were genotyped as described in Chapter 2, but using 27 cycles instead of 33 to reduce PCR artefacts. The positions of the microsatellite markers were checked in the Ensembl (July 2003) and Celera databases (July 2003).

Genotyping of single nucleotide polymorphisms (SNPs)

Fine-mapping of the candidate region by microsatellite markers resulted in a region of interest of ~450 kb between 16.80 to 17.25 Mb (based on the Ensembl map of July 2003). Fifty-two SNPs covering this region were selected from the Ensembl database. SNPs were positioned with an average spacing of 10 kb and those with a minor allele frequency >10% were preferred. These SNPs were genotyped on the patient and control pools by SNaPshot technology (AppliedBiosystems). Fifteen SNPs were subjected to individual genotyping using Assays-on-Demand or Assays-by-Design on the TaqMan 7900HT system (AppliedBiosystems). These SNPs were selected from the Celera map and five of them had also been genotyped on the pools. The positions of all SNPs were checked in the Ensembl (July 2003) and Celera databases (July 2003).

Table 1. Microsatellite markers for association analysis in DNA pools.

	Marker[a]	Position (Mb)[c]
1	D19S714*	15.57
2	D19S411	15.75
3	D19S711	15.94
4	*D19S885*	*16.05*
5	*D19S917*	*16.21*
6	*D19S1171*	*16.74*
7	*D19S930*	*16.81*
8	*D19S899**	*17.08*
9	*D19S593*	*17.15*
10	*D19S410*	*17.24*
11	*D19S429*	*17.48*
12	*D19S915*	*17.76*
13	*D4S2293*[b]	*17.92*
14	*D19S460**	*18.22*
15	*D19S898*	*18.32*
16	*D19S48*	*18.99*
17	*D19S566*	*19.01*
18	D19S603	19.21
19	D19S407*	19.91
20	D19S911	20.69
21	D19S925	21.14
22	D19S215*	21.45
23	D19S560	23.69

[a] Markers in *italics* are located within the maxLOD-1 support interval. Markers marked by an asterisk were also included in the linkage analysis (Chapter 2).

[b] Marker D4S2293 is now mapped to chromosome 19.

[c] The positions of the markers are based on the Ensembl Human Genome Map, July 2003. They were verified in the Celera Human Genome Map, July 2003, in which all positions were 0.06 Mb further.

Statistical analysis

Allele frequency estimation in DNA pools

Allele frequencies of the microsatellite markers were estimated in the patient and control pools after correction for PCR-induced stutter artefacts (Schnack et al., manuscript submitted). In short, a marker specific stutter model was derived from genotype patterns of eleven individual samples, and pool patterns were corrected for stutter using this model. Allele frequencies were then estimated from the stutter-corrected pool patterns by adding the heights of all peaks and determining the percentage that each allele contributes to the total height. SNP allele frequencies were estimated directly from the peak heights of both alleles from the DNA pool patterns.

Association analysis

The distribution of microsatellite marker alleles was compared between the patient and the control pools by χ^2-analysis. Overall P-values were calculated by testing alleles with a frequency >5% and the remaining alleles pooled into one group, resulting in a χ^2-test with n-1 degrees of freedom (df, n = number of tested alleles). Also, a P-value for the most significantly increased allele in cases was determined by testing one allele against the rest (χ^2-test with 1 df). Allele frequencies of the SNPs were also compared between the patient and control pools by a χ^2-test with 1 df. Markers and SNPs that were individually genotyped were tested by the same procedure. Haplotypes were estimated using the SNPHAP program from the individually generated genotypes (D. Clayton; available from http://www-gene.cimr.cam.ac.uk/clayton/software/).

Nine markers were also typed in the parent-case trios and analysed by the transmission/disequilibrium test (TDT). All alleles with a frequency >5% were tested for non-random transmission by χ^2-analysis with 1 df.

Results

Fine-mapping with microsatellite markers

Allele frequencies of the 23 microsatellite markers were estimated in the patient and the control pools (Table 1). Association analysis was performed for each marker by calculating overall P-values (Figure 1a) and a P-value for the most significantly increased allele in the cases (Figure 1b). Overall P-values of eleven markers were significant ($P < 0.05$, $-\log10(P) > 1.3$). It should be noted that the highly significant results of markers D19S714 and D19S930 (markers 1 and 7 from Table 1 and Figure 1) were mainly due to an increase in controls of alleles with a frequency <5% that were pooled into the rest group. When looking at the most significantly increased allele in cases, ten markers showed significant association. On comparing these results, there were two consistent regions with multiple associated markers. The first block of ~450 kb ranges from D19S930 to D19S410 (markers 7-10) and includes D19S899, the marker at which the multipoint lod score peaked (see also Chapter 2).[2] The second block lies just outside the support interval and ranges from D19S566 to D19S407 (markers 17-19). The observed

association in this block is mainly due to one specific allele with a higher frequency in cases and there is little support from overall P-values.

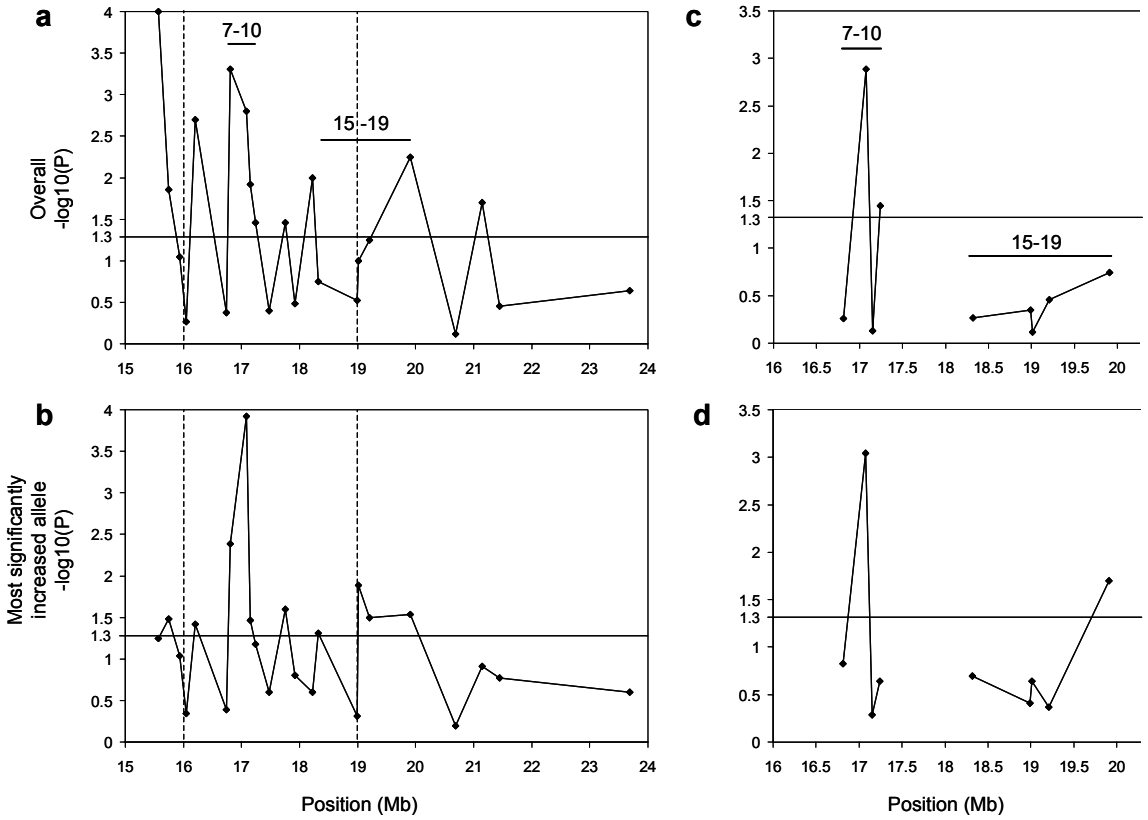


Figure 1. Association analysis of the microsatellite markers in cases and controls by DNA pooling (**a + b**) and by individual genotyping (**c + d**). P-values are presented as –log10(P) values. A P-value <0.05 was considered significant, corresponding to –log10(P) >1.3, and this threshold is indicated by the horizontal line. Positions are according to the Ensembl Human Genome Map, July 2003. The diamond symbols represent the markers, an overview of the markers is listed in Table 1. **a + c.** Overall P-value of the marker. The small horizontal lines indicate the markers that were subjected to individual genotyping with the numbers corresponding to those in Table 1. **b + d.** P-value of the most significantly increased allele in cases.

To confirm the results obtained from the DNA pools, nine markers were subjected to individual genotyping: the four markers in the first block (markers 7-10), the three markers in the second block (markers 17-19), D19S48 (marker 16) because it is very close to D19S566, and D19S898 (marker 15) because of its significant result in the pools. Association of D19S899 (marker 8) could be confirmed with an overall P-value of 0.0013 and an allele specific P-value of 0.0009 (Figures 1c and 1d). The overall P-value of

D19S410 (marker 10, P = 0.036) and the allele specific P-value of D19S407 (marker 19, P = 0.02) were also significant. Association of the other markers could not be confirmed.

The same nine markers were also genotyped in the parent-case trios, and only markers D19S48 (marker 16, P = 0.008) and D19S566 (marker 17, P = 0.006) showed preferential transmission of one specific allele. These alleles also form a haplotype, which was significantly more often transmitted (P = 0.009).

Even though the observed association of D19S899 could not be confirmed in the parent-case trios, this remains the prime region of interest. This marker is most significantly associated with coeliac disease in the case-control cohort and is located exactly at the position of the maximum lod score in the multipoint analysis. We therefore continued our search for the 19p13.1 coeliac disease gene in the region surrounding D19S899. However, the D19S48-D19S566 region is also of considerate interest and should also be subjected to further research.

Table 2. Genes in the region of interest at 19p13.12.

No[a]	Name	Function	Start position[b]	End position[b]	Ensembl ID
1	SIN3B	Unknown	16785441	16836373	ENSG00000127511
2	F2RL3	Coagulation factor II (thrombin) receptor-like 3	16845208	16847015	ENSG00000127533
3	NM_015692	Alpha-2 macroglobulin family protein VIP	16849128	16982661	ENSG00000160111
4	Q9BT25	Unknown	17005746	17031439	ENSG00000131351
5	MYO9B	Myosin IXB	17057737	17169313	ENSG00000099331
6	NM_018467	Uncharacterised hemato-poietic stem/progenitor cells protein MSD032	17171408	17175847	ENSG00000053501
7	NM_024578	Unknown	17182222	17185233	ENSG00000099330
8	NR2F6	Orphan nuclear receptor EAR-2	17187901	17201364	ENSG00000160113
9	NM_031941	AIE-75 binding protein	17206243	17220317	ENSG00000130307
10	NM_014173	Unknown	17223442	17235355	ENSG00000105393
11	NM_152363	Unknown	17239538	17243456	ENSG00000160117
12	NM_024527	Unknown	17248150	17259443	ENSG00000127220

[a] These numbers refer to the numbering of genes in Figure 2.
[b] Positions were based on the Ensembl Human Genome Map, July 2003.

SNP genotyping in region of interest

The region subjected to further study ranged from 16.80 to 17.25 Mb and contained 12 genes (Figure 2 and Table 2). None of these genes were obvious functional candidates, so they were all investigated further. Fifty-two SNPs in this region of interest, evenly distributed with an average spacing of 10 kb, were genotyped in the patient and control pools. However, seven of these SNPs turned out not to be polymorphic and eleven SNPs could not be genotyped due to PCR problems. Eleven of the remaining 34 SNPs that could be successfully analysed showed association with coeliac disease ($P < 0.05$, Figure 2). Four of them were subsequent SNPs in a 27 kb region of the same gene, the gene encoding myosin IXB (*MYO9B*). The associated microsatellite marker D19S899 is also located in this gene. The myosin IXB gene region therefore received the highest priority and fifteen SNPs in this region were subjected to individual genotyping (Figure 2). Six of them showed significant association, five of which are located in the 3' half of *MYO9B* (P-values varying from 0.02 to 0.003). The haplotype formed by the increased alleles of these five SNPs in cases is also significantly increased on comparing cases (38.8%) and controls (30.9%) ($P = 0.02$).

Discussion

Linkage analysis in Dutch affected sibpairs resulted in localisation of a susceptibility gene for coeliac disease to chromosome region 19p13.1. The candidate region spanned 3 Mb and contained 92 genes. Since fine-mapping of a region of this size would require a considerable amount of individual genotyping, we attempted to reduce the 19p13.1 candidate region by screening for microsatellite marker association in pooled DNA. Allele frequencies were compared using two strategies, as previously suggested.[4] Comparison of overall allele distributions is expected to be a sensitive measure when multiple alleles of one marker are associated with the disease. Single associated alleles are more likely to be detected by testing each allele against the combined rest. Our aim was to reduce the candidate region to a size that could be followed up by individual genotyping. Therefore, no correction was applied for the number of tested markers and alleles in order to minimize the chance of false-negative results, thereby accepting an increased false-positive rate.

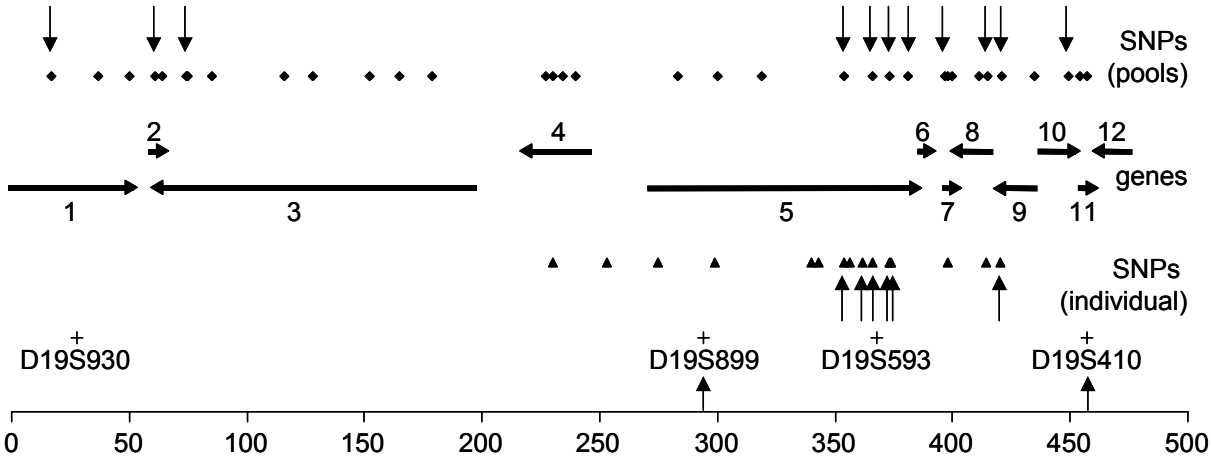


Figure 2. Overview of genes, SNPs and microsatellite markers in the region of interest on 19p13.12. Positions of genes, SNPs and markers are based on the Ensembl Human Genome Map, July 2003. The position on the X-axis is given in kb, position 0.0 corresponds to 16785441. The horizontal bars depict the genes in this region with the arrowheads indicating the transcription direction. The numbers above and below the genes correspond to the numbering in Table 2. The diamond symbols denote the 34 SNPs that were analysed in the DNA pools. The arrows indicate SNPs with significant different allele frequencies between the patient and control pools ($P < 0.05$). The triangular symbols denote the 15 SNPs that were subjected to individual genotyping and the arrows indicate the significantly associated SNPs ($P < 0.05$). The cross symbols denote the microsatellite markers that were also subjected to individual genotyping with the arrows indicating significantly associated markers ($P_{overall} < 0.05$, see Figure 1).

When looking at overall P-values obtained from the DNA pools, as many as half the markers showed significant association with coeliac disease. For some of these markers, most contribution came from the rare alleles that were pooled into one group, and that were increased in the controls. Testing for the most significantly increased allele in cases did not decrease the number of associated markers. However, significance levels were lower and only one region clearly showed the strongest association: the region around marker D19S899 (see Figures 1a and b). Association of two markers in this region, D19S899 and D19S410, could be confirmed by individual genotyping. D19S593, located between these two markers, did not show association, but this marker was not very informative, with a cumulative frequency of 80% for only two alleles. When combining all the results, the region of interest was narrowed-down to ~450 kb at 19p13.12. This region contains only 12 genes, a reduction of 80 possible candidate genes.

Genotyping of SNPs from the 450 kb region of interest in the DNA pools showed eight significantly associated SNPs in the 3' 150 kb part of this region. Individual genotyping revealed the presence of six significantly associated SNPs in a region of 70 kb. However, no SNPs have yet been subjected to individual genotyping in the final 3' 75 kb part of the region, and the disease-associated region may therefore span the entire 3' 150 kb part of the region. This region contains eight genes, four of which are of unknown function and four with functions that are not obviously compatible with coeliac disease. At this moment, the most interesting candidate gene is *MYO9B* since five of the six significantly associated SNPs are located in the 3' part of this gene, which also form a significantly increased haplotype in cases. The *MYO9B* gene encodes an unconventional myosin which exhibits a GTPase activating protein (GAP) domain for Rho family small G-proteins in its tail.[5] Myosin IXB regulates Rho activity by converting active Rho-GTP into the inactive Rho-GDP state. Rho activation results in formation of actin filament bundles (stress fibres) and focal adhesion complexes.[6] Overexpression of the rat orthologue of myosin IXB (myr5) resulted in loss of actin filaments and cell contacts by inactivation of Rho.[7] Furthermore, myosin IXB was shown to be an active motor, capable of binding to actin filaments independent of ATP presence.[8] Human myosin IXB is most highly expressed in blood leukocytes, but is also expressed at lower levels in several other tissues, including small intestine.[5] At first sight myosin IXB is not an obvious candidate for causing coeliac disease. However, it is possible that the presence of an intact cytoskeleton and/or focal adhesion complexes in the gut is necessary to deal with the gluten-derived peptides. The signal transduction capacity of myosin IXB may regulate this process.

Although most evidence points towards the *MYO9B* gene region harbouring the gene predisposing to coeliac disease, other genes cannot be excluded yet. Markers D19S48 and D19S566, located approximately 2 Mb downstream of the *MYO9B* region, showed significant increased transmission to coeliac disease patients in the parent-case trios. These results could not be confirmed in the case-control cohort. The associated allele of D19S48 has a frequency of 66% in the controls, and this high frequency may have prohibited detection of association. However, the associated allele of D19S566 has a frequency of only 17% in controls. The case-control cohort does not have sufficient power to detect allele frequency differences ≤5%, which may explain the discrepancy. On

the other hand, about 30% parents will be heterozygous for the associated allele and, therefore, enough informative transmissions to affected children in the parent-case trios can be included to obtain statistical significance. D19S48 and D19S566 are both located within a gene with unknown function (NM_033415). Four SNPs in this gene were genotyped in the DNA pools, but there was no evidence for association (data not shown). However, it cannot be excluded that this gene, or another gene residing in this region, is also predisposing to coeliac disease.

In conclusion, systematic fine-mapping of the 19p13.1 candidate region reduced the number of possible candidate genes from 92 to only eight. Additional SNPs in this region have to be tested by individual genotyping in the case-control cohort to determine the extent of the associated region. SNPs that show significant association should be confirmed in the parent-case trios and also typed in the 82 families from the linkage study to determine whether these SNPs explain the linkage. Genes within this region will have to be subjected to functional assays to determine which one is actually involved in coeliac disease.

Acknowledgements

Jackie Senior is gratefully acknowledged for improving this chapter. This study was financially supported by the Netherlands Digestive Diseases Foundation (grant WS 97-44).

References

1. Popat S, Hearle N, Hogberg L*, et al.* Variation in the CTLA4/CD28 gene region confers an increased risk of coeliac disease. *Ann Hum Genet* 2002; **66:** 125-37.
2. van Belzen MJ, Meijer JW, Sandkuijl LA*, et al.* A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* 2003; **125:** 1032-41.
3. Sham P, Bader JS, Craig I, O'Donovan M, Owen M. DNA Pooling: a tool for large-scale association studies. *Nat Rev Genet* 2002; **3:** 862-71.
4. Simonic I, Gericke GS, Ott J, Weber JL. Identification of genetic markers associated with Gilles de la Tourette syndrome in an Afrikaner population. *Am J Hum Genet* 1998; **63:** 839-46.
5. Wirth JA, Jensen KA, Post PL, Bement WM, Mooseker MS. Human myosin-IXb, an unconventional myosin with a chimerin-like rho/rac GTPase-activating protein domain in its tail. *J Cell Sci* 1996; **109 (Pt 3):** 653-61.
6. Hall A. Rho GTPases and the actin cytoskeleton. *Science* 1998; **279:** 509-14.

7. Muller RT, Honnert U, Reinhard J, Bahler M. The rat myosin myr 5 is a GTPase-activating protein for Rho in vivo: essential role of arginine 1695. *Mol Biol Cell* 1997; **8:** 2039-53.
8. Post PL, Bokoch GM, Mooseker MS. Human myosin-IXb is a mechanochemically active motor and a GAP for rho. *J Cell Sci* 1998; **111 (Pt 7):** 941-50.

Chapter 5

The *TGM2* gene and coeliac disease

The tissue transglutaminase gene is not a primary factor predisposing to coeliac disease

M.J. van Belzen, C.J.J. Mulder, P.L. Pearson, R.H.J. Houwen and C. Wijmenga

American Journal of Gastroenterology 2001; 12: 3337-3340.

Summary

Objectives: The aim of this study was to determine whether the tissue transglutaminase gene (*TGM2*) is a causal factor in the pathogenesis of coeliac disease. *Methods:* A total of 147 Dutch families with at least one biopsy-proven coeliac disease patient were available for this study. All the patients were diagnosed according to the revised European Society for Paediatric Gastroenterology and Nutrition (ESPGAN) criteria. A microsatellite marker in a non-coding region of the *TGM2* gene was investigated for both linkage and association. Linkage was tested by determining the amount of allele sharing between affected brothers and sisters (affected sibpair analysis). Association was determined by comparing transmission of certain *TGM2* alleles from parents to coeliac disease patients to the non-transmitted alleles by the transmission/disequilibrium test (TDT). *Results:* Linkage analysis did not show co-segregation of the *TGM2* gene with coeliac disease in our families, neither was there any association between certain *TGM2* alleles and coeliac disease. Furthermore, the *TGM2* gene could be excluded as a coeliac disease susceptibility gene. *Conclusions:* Our results indicate that the *TGM2* gene can be excluded as a major primary genetic factor in coeliac disease pathogenesis.

Introduction

Coeliac disease is a common food-intolerance in humans affecting 1/150 to 1/300 individuals in the Netherlands.[1,2] The disease is caused by dietary gluten, which are present in wheat, barley and rye. Ingestion of gluten leads to villous atrophy, crypt cell hyperplasia and infiltration of intestinal epithelium by lymphocytes in coeliac disease patients.[3] Coeliac disease is strongly associated with human leukocyte antigen (HLA) DQ2: 90-95% of Caucasian patients express this molecule. The HLA-DQ2 negative patients are HLA-DQ8.[4] HLA-DQ2 occurs frequently in the Dutch population, with a prevalence of ~25%.[5] It is therefore anticipated that other genes will also be involved in coeliac disease aetiology.

Gluten-derived peptides have been shown to bind to HLA-DQ2 and HLA-DQ8. These complexes are recognized by small intestinal T cell clones from coeliac disease patients. However, treatment with tissue transglutaminase (tTG) results in deamidation of specific glutamine residues in gliadin peptides. It has been shown that these modified peptides bind much better to HLA-DQ2 or HLA-DQ8 and the activity of the T cell

clones was greatly enhanced.[6,7] Interestingly, autoantibodies to tTG are present in the serum of coeliac disease patients and these are very specific markers of coeliac disease.[8] Besides deamidation of glutamine residues to glutamic acid, tTG is an enzyme that catalyses the cross-linking of proteins.[9] It is expressed in many tissues, including the small intestinal wall.[10] In coeliac disease, tTG activity in small intestinal biopsy specimens from treated and untreated patients is higher than in control samples.[11] It is unclear, however, whether this increased tTG activity is involved in the pathogenesis of coeliac disease, or whether this is a secondary effect due to the presence of the coeliac lesion.

One could hypothesize that coeliac disease patients express a certain isoform of the gene encoding tTG (*TGM2*), resulting in an inadequate response to gluten ingestion. A recent study revealed no differences in the coding sequence of the *TGM2* gene between eight patients and four controls.[12] However, a pathogenic role for the *TGM2* gene in coeliac disease should not yet be excluded. Polymorphisms in regulatory, non-coding sequences can lead to differences in the expression of the gene, and therefore in activity of the tTG protein. We studied the involvement of the *TGM2* gene in coeliac disease by using a dinucleotide repeat polymorphism in the promoter region of the gene as a marker. This marker in *TGM2* was tested for both linkage and association in a cohort of 147 Dutch families with at least one patient with coeliac disease, assuming that if a polymorphism in the *TGM2* gene causing coeliac disease exists, it will be in linkage disequilibrium (LD) with the marker.

Materials and Methods

Subjects

A total of 147 Dutch coeliac disease families were ascertained for this study. For the linkage analysis, 72 Dutch families with at least 2 children with coeliac disease were ascertained. DNA was collected from the patients and their first-degree relatives. These families contained a total of 81 affected sibpairs, which were used for affected sibpair analysis. In addition, association was studied using the transmission/disequilibrium test (TDT), which scores transmission of certain alleles to affected children. For this test, families with at least one child with coeliac disease and both parents willing to participate were collected. Forty-four of the families that participated in the linkage study were included in the association study. Another 75 Dutch families were ascertained for this

study, so that in total 119 families were available to perform the TDT test. All patients were diagnosed according to the revised European Society for Paediatric Gastroenterology and Nutrition (ESPGAN) criteria.[13] The patients in this study included both children and adults. A total of 89 healthy Dutch Caucasian controls were available to determine allele frequencies of the *TGM2* repeat polymorphism.

Analysis of the tissue transglutaminase gene

A GT repeat in the promoter region of the human *TGM2* gene on chromosome 20, 500bp upstream of the translation start site[14] (GenBank accession number U13920), was amplified by PCR using the following primers: TGM2for, 5'-ATC CAT GTC ACT GTG TCT GC-3'; TGM2rev, 5'-ATA CAG ATA CAC GCA GCA CC-3'. The size of the products was determined as described elsewhere.[15] The allele frequencies of this polymorphism were determined in 89 healthy Dutch controls. The CEPH reference samples 1331-01, 1331-02 and 1347-02 were included in all runs as control samples to allow adjustment of the sizes of the patient and control alleles between different experiments.

Statistical analysis

A linkage analysis was performed to determine whether the *TGM2* alleles segregated with coeliac disease in families. This was done by calculating maximum likelihood scores (MLS) using the Mapmaker/sibs program,[16] which calculates the allele sharing between affected siblings, and determines whether this is significantly different from the average sharing of 50% between brothers and sisters. Exclusion analysis was performed to determine whether the *TGM2* gene could be excluded as an inherited gene for coeliac disease with different relative risks. The threshold for exclusion is at a lod score ≤ -2.

To determine whether the one of the *TGM2* alleles was associated with coeliac disease in our patients, an association analysis was performed by TDT,[17] using the asstdt routine from the GAS package version 2.0 (A. Young, available from: ftp://ftp.ox.ac.uk/pub/users/ayoung). This test scores the transmission of alleles from heterozygous parents to affected children by comparing the transmitted alleles to the non-transmitted alleles. If an allele is not associated with the disease, a random transmission of 50% would be expected. Families with both parents available that were

used for the linkage analysis were also included in the association analysis. This can be justified, as it is possible that different *TGM2* alleles segregate with coeliac disease in different families. In this situation, the polymorphism studied is probably not the causal variant, but in LD with the true functional variant causing the disease. Therefore, a certain polymorphism can be linked to a disease, but not be associated with that disease. In families with more than one affected child, transmissions to all affected children were included, which is allowed when linkage, as demonstrated by affected sibpair analysis, is absent. Association analyses was only performed with alleles with a frequency of 5% or more in the population to minimize multiple testing.

Results

The GT repeat in the promoter region of the human *TGM2* gene was first tested in 178 chromosomes from subjects without coeliac disease. Six different alleles were detected, ranging in size from 298 bp to 308 bp (Table 1). The observed heterozygosity of this microsatellite marker in the Dutch population was 67.6%. Alleles 2 and 4 were frequent in our population, but alleles 1, 3 and 5 occurred much less frequently. Allele 6 was present in only one person. The polymorphic repeat was investigated in the coeliac disease families. All alleles, except allele 6, were present in this group. However, a new allele of 296 bp was present on one chromosome in one person.

Table 1. Frequency of the *TGM2* promoter polymorphism alleles in the Dutch population.

Allele	Size (bp)[a]	Allele frequencies
1	298	0.13
2	300	0.44
3	302	0.06
4	304	0.33
5	306	0.03
6	308	0.01

[a] Allele sizes of the CEPH reference samples: 1331-01, 300/300 bp; 1331-02, 300/300 bp and 1347-02, 300/304 bp.

Linkage analysis in 81 affected sibpairs demonstrated a random distribution of *TGM2* alleles, with 25% of the affected sibpairs sharing 0 alleles, 50% sharing 1 allele and 25% sharing 2 alleles. This allele distribution resulted in a maximum likelihood score (MLS) of 0.0, as expected. The results of the exclusion analysis showed that the *TGM2*

gene could be excluded as a coeliac disease locus with a relative risk of 2 or higher (Table 2). Therefore, the *TGM2* gene could be excluded as a major coeliac disease locus.

Table 2. Results of exclusion analysis.

Relative risk	Lod score
1.25	-0.48
1.5	-1.08
1.75	-1.68
2	-2.24
2.5	-3.23
3	-4.08
3.5	-4.80
4	-5.44

Association analysis of the four alleles with a frequency of 5% or more (i.e. alleles 1, 2, 3 and 4) was performed in 119 families. If the *TGM2* gene is involved in coeliac disease, the *TGM2* allele that confers risk to coeliac disease should be transmitted significantly more often to affected children than not transmitted. The presence of a *TGM2* risk allele was tested by performing a TDT test (Table 3). Our results clearly show that none of these alleles were preferentially transmitted to coeliac disease patients.

Table 3. Transmission of the major *TGM2* alleles to coeliac disease patients.

Allele	Transmitted	Not transmitted	P-value
1	41	47	NS[a]
2	85	80	NS
3	17	22	NS
4	70	62	NS

[a] NS: not significant.

Discussion

The major genetic contribution to coeliac disease known to date comes from the DQA1 and DQB1 genes in the HLA-region, encoding the HLA-DQ heterodimer. Ninety-five percent of coeliac disease patients express the HLA-DQ2 protein. However, since 25% of the general population also express HLA-DQ2, it is likely that other genes are also involved in coeliac disease. This can also be concluded from the results of studies that tried to determine the contribution of the HLA-region.[18,19] Both studies concluded that the relative risk for a sibling of a coeliac disease patient (λ_s) contributed by the HLA-locus could not be more than 4. In view of the total predicted λ_s for coeliac disease of

approximately 25, which is based on a population prevalence of 1 in 250 and a 10% risk for a sibling, non-HLA loci must play a considerable role in coeliac disease.

However, none of the two whole genome screens that have been performed so far[20,21] have identified loci with a lod score higher than the HLA-region, except for one locus in the Irish population on the short arm of chromosome 6, telomeric to HLA.[20] However, this result was not replicated by other groups.[21-23] So, it seems that there may be many additional loci, which each contribute only a little to the total genetic risk.

An alternative to the whole genome approach is the study of candidate genes based on knowledge of the disease process. The gene encoding tissue transglutaminase can be considered as an attractive functional candidate gene for coeliac disease. First, tTG is involved in the repair of damaged tissue.[9] In the small intestinal mucosa of coeliac disease patients, constant inflammation and tissue damage is present. Second, it is capable of deamidating gluten peptides, thereby creating a strong epitope for HLA-DQ2 and HLA-DQ8 proteins.[6,7] Third, its activity is increased in the gut of both untreated coeliac disease patients and patients on a gluten-free diet.[11] In particular, the increased activity in treated patients points to the possible effect of a functional polymorphism in the *TGM2* gene itself. It is unknown however whether expression of tTG in the gut of coeliac disease patients is also increased. And fourth, the tTG enzyme was identified as the target of the endomysial antibodies,[8] although the exact role of these antibodies in the disease process remains unclear.[24]

To implicate tTG as a causal factor in the pathogenesis of coeliac disease, a functional polymorphism in the *TGM2* gene should be involved in the pathogenesis of the disease. The repeat in the promoter region of the *TGM2* gene turned out to be highly polymorphic and proved to be a good marker for studying the involvement of this gene in coeliac disease. It is not necessary however that this repeat polymorphism itself will affect the activity or expression of the protein. If a functional variant, either coding or non-coding, causing coeliac disease is present in the *TGM2* gene, it is assumed that the alleles of the repeat polymorphism will be in LD with this functional variant in coeliac disease families. Therefore, the alleles of the repeat polymorphism will segregate with the disease in coeliac disease families and, as a consequence, increased sharing of the alleles will be present in affected siblings, which can be detected by linkage analysis.

Furthermore, if the functional variant is in LD with the same repeat allele in unrelated patients, than this specific repeat allele would also be associated with coeliac disease.

The polymorphic repeat in the *TGM2* gene was used for linkage and association studies. Affected sibpair analysis showed that the alleles of this marker did not co-segregate with the coeliac disease in our families. Furthermore, the *TGM2* gene could be excluded as a susceptibility locus for coeliac disease with a relative risk (λ_s) of 2 or higher. These results were confirmed by association analysis by TDT, which demonstrated that none of the *TGM2* alleles were associated with coeliac disease.

This is the first report of a genetic approach in a large number of families to study whether variants in the *TGM2* gene are inherited factors in coeliac disease. Our findings clearly show that it is very unlikely that polymorphisms in the *TGM2* gene are causally related to coeliac disease. These results complement those of Aldersley et al.,[12] who found no differences in the coding region of the *TGM2* gene between a small number of coeliac disease patients and controls. The increased tTG activity that is measured in small intestinal biopsy samples of coeliac disease patients could be a secondary effect of the disease process, even in patients on a gluten-free diet. However, it can also be caused by polymorphisms in genes coding for proteins that regulate transcription of the *TGM2* gene. To find such other genes, genome-wide scans in multiply affected families have to be performed. Further understanding of the role of tTG in coeliac disease pathogenesis will have to be elucidated by future studies.

Acknowledgements

The authors wish to thank Eric Strengman and Steven Bakker for their help in genotyping of the *TGM2* repeat polymorphism. We are grateful to Arjenne Kors, Anath Oren and Winny van der Hoeven for collecting the coeliac disease families and to Jackie Senior for improving the manuscript. This study was supported by the Dutch Digestive Diseases Foundation, grant WS 97-44.

References

1. Csizmadia CG, Mearin ML, von Blomberg BM, Brand R, Verloove-Vanhorick SP. An iceberg of childhood coeliac disease in the Netherlands. *Lancet* 1999; **353:** 813-4.

2. Rostami K, Mulder CJ, Werre JM*, et al.* High prevalence of celiac disease in apparently healthy blood donors suggests a high prevalence of undiagnosed celiac disease in the Dutch population. *Scand J Gastroenterol* 1999; **34:** 276-9.
3. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 1992; **102:** 330-54.
4. Sollid LM, Thorsby E. HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 1993; **105:** 910-22.
5. Moghaddam PH, de Knijff P, Schipper RF*, et al.* Selective co-evolution of the D6STNFa microsatellite region with HLA class I and II loci. *Tissue Antigens* 1998; **52:** 213-9.
6. Sjostrom H, Lundin KE, Molberg O*, et al.* Identification of a gliadin T-cell epitope in coeliac disease: general importance of gliadin deamidation for intestinal T-cell recognition. *Scand J Immunol* 1998; **48:** 111-5.
7. van de Wal Y, Kooy YM, van Veelen PA*, et al.* Small intestinal T cells of celiac disease patients recognize a natural pepsin fragment of gliadin. *Proc Natl Acad Sci U S A* 1998; **95:** 10050-4.
8. Dieterich W, Ehnis T, Bauer M*, et al.* Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nat Med* 1997; **3:** 797-801.
9. Greenberg CS, Birckbichler PJ, Rice RH. Transglutaminases: multifunctional cross-linking enzymes that stabilize tissues. *Faseb J* 1991; **5:** 3071-7.
10. Thomazy V, Fesus L. Differential expression of tissue transglutaminase in human cells. An immunohistochemical study. *Cell Tissue Res* 1989; **255:** 215-24.
11. Bruce SE, Bjarnason I, Peters TJ. Human jejunal transglutaminase: demonstration of activity, enzyme kinetics and substrate specificity with special relation to gliadin and coeliac disease. *Clin Sci (Lond)* 1985; **68:** 573-9.
12. Aldersley MA, Hamlin PJ, Jones PF, Markham AF, Robinson PA, Howdle PD. No polymorphism in the tissue transglutaminase gene detected in coeliac disease patients. *Scand J Gastroenterol* 2000; **35:** 61-3.
13. Walker-Smith JA. Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. *Arch Dis Child* 1990; **65:** 909-11.
14. Lu S, Saydak M, Gentile V, Stein JP, Davies PJ. Isolation and characterization of the human tissue transglutaminase gene promoter. *J Biol Chem* 1995; **270:** 9748-56.
15. Wijmenga C, van den Heuvel LP, Strengman E*, et al.* Localization of the ICF syndrome to chromosome 20 by homozygosity mapping. *Am J Hum Genet* 1998; **63:** 803-9.
16. Kruglyak L, Lander ES. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 1995; **57:** 439-54.
17. Spielman RS, Ewens WJ. The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 1996; **59:** 983-9.
18. Bevan S, Popat S, Braegger CP*, et al.* Contribution of the MHC region to the familial risk of coeliac disease. *J Med Genet* 1999; **36:** 687-90.
19. Houlston RS, Ford D. Genetics of coeliac disease. *Qjm* 1996; **89:** 737-43.
20. Zhong F, McCombs CC, Olson JM*, et al.* An autosomal screen for genes that predispose to celiac disease in the western counties of Ireland. *Nat Genet* 1996; **14:** 329-33.
21. Greco L, Corazza G, Babron MC*, et al.* Genome search in celiac disease. *Am J Hum Genet* 1998; **62:** 669-75.
22. Brett PM, Yiannakou JY, Morris MA*, et al.* A pedigree-based linkage study of coeliac disease: failure to replicate previous positive findings. *Ann Hum Genet* 1998; **62 (Pt 1):** 25-32.
23. Houlston RS, Tomlinson IP, Ford D*, et al.* Linkage analysis of candidate regions for coeliac disease genes. *Hum Mol Genet* 1997; **6:** 1335-9.

24. Molberg O, McAdam SN, Sollid LM. Role of tissue transglutaminase in celiac disease. *J Pediatr Gastroenterol Nutr* 2000; **30:** 232-40.

Chapter 6

The *CTLA4* gene and coeliac disease

The *CTLA4* +49A/G polymorphism in Dutch coeliac disease patients

M.J. van Belzen, C.J.J. Mulder, P.L. Pearson, R.H.J. Houwen and C. Wijmenga

Submitted for publication.

Summary

Coeliac disease is an autoimmune disorder, characterised by villous atrophy of the small intestine, which results from a T-cell mediated response to gluten-derived peptides. The cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) is involved in regulation of T-cell activation and the *CTLA4* +49 A/G polymorphism has been implicated in several autoimmune disorders. Association of this polymorphism with coeliac disease has been demonstrated in some populations, but not in others. The role of the +49 A/G variant in coeliac disease patients of Dutch origin was determined by typing the polymorphism in 210 coeliac disease patients and 208 controls. Neither allele nor genotype frequencies were significantly different distributed between cases and controls. However, a small effect of the GG-genotype was observed, which was present in 18% of cases and in 11% of controls (OR 1.78, 95% CI 0.99-3.31, p = 0.038). Although these findings suggest that the *CTLA4* gene is not a major determinant of coeliac disease susceptibility in the Dutch population, the GG-genotype may confer an increased risk for developing the disease. Recently it was shown that the haplotype carrying the G-allele is associated with lower mRNA levels of the soluble CTLA-4 isoform, providing a possible mechanism for the T-cell mediated destruction of the small intestine.

Introduction

Coeliac disease is an autoimmune disorder and is strongly associated with the HLA region. The majority of the patients express the HLA-DQ2 protein and almost all remaining patients express HLA-DQ8. Coeliac disease is characterised by an inappropriate T-cell response to gluten in the small intestine. Gluten proteins are present in wheat, barley and rye, and gluten-derived peptides can bind to HLA-DQ2 and DQ8 on the surface of antigen-presenting cells. These complexes are recognised by gluten-specific T-cells in the gut of coeliac disease patients, resulting in inflammation, crypt hyperplasia and villous atrophy.[1,2] The HLA association only partly explains the genetic contribution, so non-HLA genes must also be involved in coeliac disease.

The cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and the cell designation 28 antigen (CD28) are involved in regulation of T-cell activation. CD28 induces the T-cell response, whereas CTLA-4 has an inhibitory effect.[3] Because of their functions they are interesting candidate genes for autoimmune disorders. The genes

encoding these proteins are both located within a 150 kb region on chromosome 2q33-34. The A/G polymorphism at position +49 in exon 1 of the *CTLA4* gene has been extensively studied in autoimmune disorders and association with type 1 diabetes mellitus,[4] Graves' disease,[5,6] Hashimoto's thyroiditis[7] and multiple sclerosis[8] has been demonstrated (for a review, see[9]). It is likely that coeliac disease shares a common genetic background with other autoimmune diseases, so the *CTLA4* gene may be involved in coeliac disease as well. We investigated the contribution of the *CTLA4* gene to coeliac disease risk by typing the +49A/G polymorphism in a Dutch case-control cohort.

Subjects and Methods

Subjects

A total of 216 independent, biopsy-proven coeliac disease patients of Dutch origin were available for this study. Their mean age was 39 years and 65% were female. The control group consisted of 216 sex- and age-matched independent Dutch individuals. The study was approved by the Medical Ethics Committee of the University Medical Centre Utrecht and written informed consent was obtained from all the participants.

Genotyping

Genotypes of the A/G polymorphism at position +49 of the *CTLA4* gene were determined by PCR-restriction fragment enzyme analysis (PCR-RFLP) with the *Fnu*4HI enzyme. The fragment containing the polymorphism was amplified with primers 5'-TTG CCT TGG ATT TCA GTG GC-3' and 5'-CTG CTG AAA CAA ATG AAA CCC-3', resulting in a fragment of 175 bp. Digestion with *Fnu*4H1 of the product containing the G-allele resulted in fragments of 42 bp and 133 bp. The genotyping procedure was verified by sequencing of three samples with predicted genotypes AA, AG and GG, respectively. These three samples were included as positive controls in each of the five sample plates.

Statistical analysis

Allele and genotype frequencies of the *CTLA4* +49 A/G polymorphism were determined in cases and controls. Significance of allele and genotype distributions was tested by means of a χ^2 test with 1 and 2 degrees of freedom (df), respectively.

Association of the GG-genotype was tested versus the other genotypes by χ^2 analysis with 1 df. The genotype frequencies were tested for Hardy-Weinberg equilibrium by χ^2 analysis, with a value of $p<0.05$ considered as not in equilibrium. Odds ratios for G-allele and the GG-genotype versus AA and AG genotypes were calculated using the Woolf-Haldane correction.

Results

The +49 A/G polymorphism in the *CTLA4* gene could be genotyped in 210 cases and 208 controls (Table 1). The genotypes in the controls were in Hardy-Weinberg equilibrium ($\chi^2 = 0.48$, $p = 0.49$). The frequency of the G-allele was slightly increased in the coeliac disease patients, but this difference was not significant. The distribution of the three genotypes was not significantly different between cases and controls. However, the GG-genotype was more prevalent in the coeliac disease patients, with a frequency of 18% in cases compared to a frequency of 11% in controls (OR 1.78, 95% CI 0.99-3.31, $p = 0.038$).

Table 1. Allele and genotype frequencies of the *CTLA4* +49A/G polymorphism.

	Coeliac disease (n = 210)	Controls (n = 208)	P-value (χ^2)	OR (95% CI)
Allele				
A	0.61	0.66		
G	0.39	0.34	0.14 (2.17)	1.23 (0.92-1.65)
Genotype[a]				
AA	0.39	0.42		
AG	0.43	0.47		
GG	0.18	0.11	0.038 (4.27)	1.78 (0.99-3.31)

[a] P-value for the global genotype distribution: P = 0.12 ($\chi^2 = 4.29$, 2 df).

Discussion

The *CTLA4* gene has been implicated as a general susceptibility gene for autoimmune diseases[9]. The +49 A/G polymorphism is the most extensively studied, as this is the only polymorphism in the *CTLA4* gene that alters an amino acid.[10] The G-allele encodes an alanine residue at codon 17 of the protein, whereas the A-allele encodes a threonine residue. Type 1 diabetes mellitus, Graves' disease, Hashimoto's thyroiditis and multiple

sclerosis have all been associated with the G-allele. So the codon 17 alanine residue may be common denominator for autoimmunity.

In our sample, the frequency of the G-allele was also found to be increased in the cases, although not significantly. Even though the genotype distributions were not significantly different between cases and controls, there was a small effect of the GG-genotype. These results indicate that the GG-genotype may be associated with a slightly increased risk for coeliac disease in the Dutch population.

Several other groups have investigated the involvement of the *CTLA4* gene in coeliac disease, but the results have been conflicting. Association of the +49 A/G polymorphism has been reported in French and Swedish/Norwegian populations,[11-13] but was absent in Italian, Tunisian and Finnish populations and in a combined sample of families from Northern Europe.[14-16] Recently, a meta-analysis of these studies was performed.[16] Combining data from the seven studies that were analysed by the transmission/disequilibrium test (TDT) resulted in a non-significant p-value of 0.20. One may conclude that this polymorphism does not contribute to coeliac disease risk, but lack of power of the TDT approach may also have influenced these results. The TDT design may not be an appropriate test for association since parents have a 55% probability of being homozygous at this locus, based on the allele frequencies in Dutch controls. This dramatically reduces the power and a case-control design may be preferred. Inclusion of the only case-control study[11] in the meta-analysis reduced the p-value to 0.019. However, the results from this case-control study should be regarded with circumspection as the genotypes of the controls were not in Hardy-Weinberg equilibrium, which may have resulted from genotyping errors.

Remarkably, the positive association of the +49 A/G polymorphism with coeliac disease in French and Swedish/Norwegian populations was with the A-allele.[11-13] This argues in favour of the possibility that not the +49 A/G polymorphism itself, but a variant in linkage disequilibrium with this polymorphism is contributing to the disease risk. Indeed, a recent publication[10] showed that the causal variant involved in Graves' disease, autoimmune hypothyroidism and type 1 diabetes mellitus is most probably located in the 6.1 kb region 3' of the *CTLA4* gene, and the +49 A/G polymorphism could be rejected as a causal polymorphism. Two very common *CTLA4* haplotypes were present: one protective haplotype with the A-allele at position +49 and one disease-

susceptible haplotype with the G-allele at position +49. More importantly, the disease-susceptible haplotype encodes an alternatively spliced CTLA-4 mRNA, resulting in lower mRNA levels of the soluble CTLA-4 isoform compared to the protective haplotype.

Taken together, the *CTLA4* gene seems to confer an increased risk to coeliac disease, at least in some populations. The risk is probably mediated by the disease-susceptible haplotype, associated with lower mRNA levels and possibly with lower protein levels of the soluble CTLA-4 isoform. This may result in a disruption in the regulation of the T-cell response by CTLA-4 and CD28, leading to a T-cell mediated destruction of the gut. Determination of the mRNA levels of the soluble CTLA-4 isoform in small intestinal tissue from coeliac disease patients may provide more insight into the role of the *CTLA4* gene in coeliac disease pathogenesis.

Acknowledgements

We thank Winny van der Hoeven and Alfons Bardoel for collecting the cases and controls, Jos Meijer for re-evaluating the initial biopsy specimens of the patients included in this study and Jackie Senior for improving the manuscript. This study was supported by a grant from the Netherlands Digestive Diseases Foundation (WS 97-44).

References

1. Sollid LM. Molecular basis of celiac disease. *Annu Rev Immunol* 2000; **18:** 53-81.
2. Schuppan D. Current concepts of celiac disease pathogenesis. *Gastroenterology* 2000; **119:** 234-42.
3. Greenfield EA, Nguyen KA, Kuchroo VK. CD28/B7 costimulation: a review. *Crit Rev Immunol* 1998; **18:** 389-418.
4. Marron MP, Raffel LJ, Garchon HJ*, et al.* Insulin-dependent diabetes mellitus (IDDM) is associated with CTLA4 polymorphisms in multiple ethnic groups. *Hum Mol Genet* 1997; **6:** 1275-82.
5. Donner H, Rau H, Walfish PG*, et al.* CTLA4 alanine-17 confers genetic susceptibility to Graves' disease and to type 1 diabetes mellitus. *J Clin Endocrinol Metab* 1997; **82:** 143-6.
6. Heward JM, Allahabadia A, Armitage M*, et al.* The development of Graves' disease and the CTLA-4 gene on chromosome 2q33. *J Clin Endocrinol Metab* 1999; **84:** 2398-401.
7. Donner H, Braun J, Seidl C*, et al.* Codon 17 polymorphism of the cytotoxic T lymphocyte antigen 4 gene in Hashimoto's thyroiditis and Addison's disease. *J Clin Endocrinol Metab* 1997; **82:** 4130-2.
8. Harbo HF, Celius EG, Vartdal F, Spurkland A. CTLA4 promoter and exon 1 dimorphisms in multiple sclerosis. *Tissue Antigens* 1999; **53:** 106-10.
9. Kristiansen OP, Larsen ZM, Pociot F. CTLA-4 in autoimmune diseases--a general susceptibility gene to autoimmunity? *Genes Immun* 2000; **1:** 170-84.

10. Ueda H, Howson JM, Esposito L*, et al.* Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 2003; **30:** 30.
11. Djilali-Saiah I, Schmitz J, Harfouch-Hammoud E, Mougenot JF, Bach JF, Caillat-Zucman S. CTLA-4 gene polymorphism is associated with predisposition to coeliac disease. *Gut* 1998; **43:** 187-9.
12. Naluai AT, Nilsson S, Samuelsson L*, et al.* 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* 2000; **56:** 350-5.
13. Popat S, Hearle N, Wixey J*, et al.* Analysis of the CTLA4 gene in Swedish coeliac disease patients. *Scand J Gastroenterol* 2002; **37:** 28-31.
14. Clot F, Fulchignoni-Lataud MC, Renoux C*, et al.* Linkage and association study of the CTLA-4 region in coeliac disease for Italian and Tunisian populations. *Tissue Antigens* 1999; **54:** 527-30.
15. Holopainen P, Arvas M, Sistonen P*, et al.* CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to celiac disease. A linkage and family-based association study. *Tissue Antigens* 1999; **53:** 470-5.
16. Popat S, Hearle N, Hogberg L*, et al.* Variation in the CTLA4/CD28 gene region confers an increased risk of coeliac disease. *Ann Hum Genet* 2002; **66:** 125-37.

Chapter 7

The *IFNG* gene and coeliac disease

The intragenic interferon-γ CA-repeat polymorphism is not associated with coeliac disease

M.J. van Belzen, A. Fariña Sarasqueta, A.F.J. Bardoel, P.L. Pearson, C.J.J. Mulder, R.H.J. Houwen and C. Wijmenga

Submitted as part of a paper on expression and genetics of IFNG in coeliac disease.

Summary

Background: Coeliac disease is a complex inflammatory disorder of the small intestine. It is characterized by an inappropriate T cell response to dietary gluten in genetically susceptible individuals. One of the most important cytokines in this process is interferon-γ (IFN-γ), which becomes highly expressed in the gut of patients when stimulated with gluten. Notably, expression of IFN-γ is also increased in treated patients with complete clinical recovery, suggesting that a genetic factor influencing the expression of the *IFNG* gene may predispose to coeliac disease. We therefore aimed to investigate whether the *IFNG* gene is a causal factor in coeliac disease pathogenesis by testing for association with a microsatellite marker in the first intron, of which a certain allele is associated with high expression of IFN-γ. *Methods:* The CA repeat in *IFNG* was genotyped in 207 coeliac disease patients and 210 normal controls. The marker was also genotyped in 122 parent-case trios. Overall association of the polymorphism was determined, as well as a specific effect of the allele associated with high expression of IFN-γ. *Results:* The allele frequencies were not significantly different distributed between cases and controls. Furthermore, there was no evidence for a specific effect of the highly expressed allele. Likewise, all alleles were randomly transmitted to affected children in the parent-case trios. *Discussion:* The *IFNG* gene is not a predisposing factor in coeliac disease, even though IFN-γ is a very important cytokine in the pathogenesis of this disease.

Introduction

Coeliac disease is a gluten-sensitive enteropathy of the small intestine, characterised by increased numbers of lymphocytes in the gut epithelium and lamina propria, crypt hyperplasia and villous atrophy. Clinical symptoms include gastrointestinal complaints such as diarrhoea and abdominal distension but also fatigue, weight loss, growth retardation, anaemia, osteopenia and failure to thrive.[1] The disease is caused by ingestion of wheat gluten and similar proteins in barley and rye. Gluten peptides are not completely digested by the enzymes in the stomach and small intestine.[2] The partially digested gluten peptides cross the gut epithelium by an unknown mechanism. These peptides are bound by HLA-DQ2 and -DQ8 proteins on the cell surface of antigen presenting cells and recognized by the T cells in the lamina propria.[3] This explains the well-known genetic

association between coeliac disease and HLA-DQ, with over 90% of patients expressing DQ2 and almost all DQ2-negative patients expressing DQ8. Withdrawal of gluten from the diet usually results in normalisation of the small intestinal lesion and complete disappearance of the clinical symptoms.

Interferon-γ (IFN-γ) plays an important role in coeliac disease pathogenesis. The expression of IFN-γ mRNA is highly increased in small intestinal biopsies from untreated patients compared to histologically normal controls. [4-7] The proportion of IFN-γ producing T cells was also increased in untreated patients.[4,5,7] Moreover, *in vitro* gliadin challenge of biopsies from treated patients (patients on a gluten-free diet with normal histology) showed IFN-γ mRNA levels that were increased to the levels detected in untreated patients.[5] In addition, cultured duodenal biopsy specimens from normal individuals showed damage to enterocytes when stimulated with IFN-γ or supernatants from gluten-sensitive T cells.[8] The toxic effects of the supernatants from gluten-sensitive T cells could be blocked by administration of anti-IFN-γ antibodies, indicating that IFN-γ is involved in the destruction of the enterocytes.

Notably, expression of IFN-γ by intra-epithelial lymphocytes was still increased in treated, symptom-free patients with normal histology.[4] One explanation could be a predisposition of coeliac disease patients to high IFN-γ expression. Polymorphisms in genes involved in the expression of IFN-γ may be associated with a high expression level of IFN-γ and, therefore, with coeliac disease. However, a polymorphism in the *IL12B* gene, encoding the p40 subunit of interleukin (IL)-12 which is a strong inducer of IFN-γ secretion, was not associated with increased risk to coeliac disease.[9,10] Likewise, no evidence has been found for association of coeliac disease with the gene encoding interferon regulatory factor (IRF)-1, a transcription factor potentially involved in expression of IL-12.[10] Therefore, the IFN-γ gene (*IFNG*) itself may be considered as a candidate gene.

The gene encoding IFN-γ (*IFNG*) is located on chromosome region 12q15 and it does not contain any polymorphisms in coding regions.[11] One of the intronic polymorphisms is a CA-repeat in the first intron of the gene, which was shown to be in complete linkage disequilibrium with an A/T single nucleotide polymorphism (SNP) at

position +874. This SNP showed an absolute correlation between the presence of a T-allele at +874 and the presence of 12 CA repeats.[12] It has been suggested that these alleles are associated with increased levels of IFN-γ production,[13-15] although others were unable to confirm this finding.[16,17] The CA-repeat polymorphism has been shown to be associated with several auto-immune disorders like type 1 diabetes mellitus,[18,19] rheumatoid arthritis,[20] and multiple sclerosis.[21,22] The aim of this study was to evaluate the role of the *IFNG* gene in coeliac disease by association analysis of the CA-repeat in two cohorts of Dutch coeliac disease patients.

Subjects and Methods

Subjects

The first cohort consisted of 207 independent Dutch biopsy-proven coeliac disease patients and 210 independent Dutch individuals as controls. The cases and controls both consisted of Caucasians with a mean age of 39 years and a male-to-female ratio of 35% vs 65%. The second cohort comprised 122 Caucasian, Dutch parents-case trios for analysis by the transmission/disequilibrium test (TDT), consisting of one child with biopsy-proven coeliac disease and both parents. The patients in this cohort had a mean age of 17 years and also 65% were females.

IFNG CA-repeat typing

PCR fragments containing the *IFNG* CA-repeat polymorphism were obtained in a reaction volume of 10 µl containing 25 ng of DNA, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 50 ng of forward primer 5'-TTA TTC TTA CAA CAC AAA ATC AAA TC-3', 50 ng of reverse, fluorescence-labelled primer 5'-ATA CAA AAA CAA AAA ACA GCA AAG C-3' and 0.4 U AmpiTaq Gold. The PCR products of 190-200 bp were separated on a 3700 DNA sequencer and analysed by Genescan 3.5 and Genotyper 2.0 software. The nomenclature of the alleles is adapted from the report by [14], in which five alleles of this polymorphism are described (alleles 1-5 correspond to 11-15 CA-repeats, respectively). In our sample, allele 1 was not present, and alleles 6 and 7 (16 and 17 CA-repeats, respectively) had frequencies below 0.5% and were not included in the analysis.

Statistical analysis

Allele and genotype frequencies were compared between cases and controls and P-values were obtained by χ^2 analysis. Alleles transmitted and non-transmitted to affected offspring in the parent-case trios were compared by a χ^2 test with 1 df.

Results and Discussion

The overall distribution of *IFNG* CA-repeat alleles was not significantly different between cases and controls (Table 1). The influence of allele 2 (12 CA-repeats) was specifically investigated, as this allele was shown to be associated with high *in vitro* expression of IFN-γ. Homozygosity or heterozygosity for this allele did not significantly contribute to disease risk (Table 1). Likewise, no significant distortion of random transmission of CA-repeat alleles to affected offspring was present in the parent-case trios (Table 2). In conclusion, there was no evidence for association of any of the *IFNG* CA-repeat alleles with coeliac disease in the two cohorts. *IFNG* is therefore unlikely to be a predisposing gene in coeliac disease.

Table 1. Allele and genotype frequencies of the *IFNG* CA-repeat polymorphism in cases and controls.

	Coeliac disease (N = 207)	Controls (N = 210)
Allele[a]		
2	0.46	0.49
3	0.44	0.41
4	0.06	0.06
5	0.04	0.04
Genotype[a]		
2,2	0.23	0.22
2,3	0.37	0.41
2,4	0.04	0.08
2,5	0.04	0.04
2,7	0.004	0.0
X,X[b]	0.32	0.25

[a] No significant differences in either allele or genotype frequencies were present between cases and controls.

[b] Allele X is not allele 2.

Table 2. Transmission of *IFNG* CA-repeat alleles to coeliac disease patients.

Allele	Transmitted[a]	Not transmitted
2	51	51
3	47	51
4	14	12
5	7	6

[a] None of the alleles were significantly over-transmitted to affected offspring.

The results obtained in this study complement those from our linkage analysis in Dutch affected sibpairs, in which we found no evidence for linkage of coeliac disease to 12q15.[23] It is unlikely that *IFNG* will make a strong contribution to coeliac disease in other populations, since no evidence for linkage to this region was found in eight other genome-wide screens.[24-31] Nevertheless, it cannot be excluded that the up-regulation of IFN-γ mRNA expression by intraepithelial lymphocytes of treated coeliac disease patients could still be due to a genetic predisposition by a polymorphism in a gene regulating the expression of *IFNG*. Alternatively, the gluten-free diet may contain trace amounts of gluten, which are not sufficient to induce histological changes, but may activate T cells in the small intestinal tissue. On the other hand, persisting T cell activation may just be a consequence of the lesions in the small intestine that were present during the active disease period.

Acknowledgements

We thank Jackie Senior for improving the manuscript. This study was financially supported by the Netherlands Digestive Diseases Foundation (grant WS 97-44).

References

1. Farrell RJ, Kelly CP. Celiac sprue. *N Engl J Med* 2002; **346:** 180-8.
2. Shan L, Molberg O, Parrot I*, et al.* Structural basis for gluten intolerance in celiac sprue. *Science* 2002; **297:** 2275-9.
3. Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* 2002; **2:** 647-55.
4. Forsberg G, Hernell O, Melgar S, Israelsson A, Hammarstrom S, Hammarstrom ML. Paradoxical coexpression of proinflammatory and down-regulatory cytokines in intestinal T cells in childhood celiac disease. *Gastroenterology* 2002; **123:** 667-78.
5. Nilsen EM, Jahnsen FL, Lundin KE*, et al.* Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease. *Gastroenterology* 1998; **115:** 551-63.

6. Lahat N, Shapiro S, Karban A, Gerstein R, Kinarty A, Lerner A. Cytokine profile in coeliac disease. *Scand J Immunol* 1999; **49:** 441-6.
7. Kontakou M, Sturgess RP, Przemioslo RT, Limb GA, Nelufer JM, Ciclitira PJ. Detection of interferon gamma mRNA in the mucosa of patients with coeliac disease by in situ hybridisation. *Gut* 1994; **35:** 1037-41.
8. Przemioslo RT, Lundin KE, Sollid LM, Nelufer J, Ciclitira PJ. Histological changes in small bowel mucosa induced by gliadin sensitive T lymphocytes can be blocked by anti-interferon gamma antibody. *Gut* 1995; **36:** 874-9.
9. Louka AS, Torinsson Naluai A, D'Alfonso S*, et al.* The IL12B gene does not confer susceptibility to coeliac disease. *Tissue Antigens* 2002; **59:** 70-2.
10. Seegers D, Borm ME, van Belzen MJ*, et al.* Lack of association between *IL12B* and *IRF1* gene polymorphisms and susceptibility to celiac disease. *Eur J Immunogenet* 2003**:** in press.
11. Bream JH, Carrington M, O'Toole S*, et al.* Polymorphisms of the human IFNG gene noncoding regions. *Immunogenetics* 2000; **51:** 50-8.
12. Pravica V, Perrey C, Stevens A, Lee JH, Hutchinson IV. A single nucleotide polymorphism in the first intron of the human IFN- gamma gene: absolute correlation with a polymorphic CA microsatellite marker of high IFN-gamma production. *Hum Immunol* 2000; **61:** 863-6.
13. Hoffmann SC, Stanley EM, Darrin Cox E*, et al.* Association of cytokine polymorphic inheritance and in vitro cytokine production in anti-CD3/CD28-stimulated peripheral blood lymphocytes. *Transplantation* 2001; **72:** 1444-50.
14. Pravica V, Asderakis A, Perrey C, Hajeer A, Sinnott PJ, Hutchinson IV. In vitro production of IFN-gamma correlates with CA repeat polymorphism in the human IFN-gamma gene. *Eur J Immunogenet* 1999; **26:** 1-3.
15. Warle MC, Farhan A, Metselaar HJ*, et al.* Are cytokine gene polymorphisms related to in vitro cytokine production profiles? *Liver Transpl* 2003; **9:** 170-81.
16. Nieters A, Brems S, Becker N. Cross-sectional study on cytokine polymorphisms, cytokine production after T-cell stimulation and clinical parameters in a random sample of a German population. *Hum Genet* 2001; **108:** 241-8.
17. Cartwright N, Demaine A, Jahromi M, Sanders H, Kaminski ER. A study of cytokine protein secretion, frequencies of cytokine expressing cells and IFN-G gene polymorphisms in normal individuals. *Transplantation* 1999; **68:** 1546-52.
18. Awata T, Matsumoto C, Urakami T, Hagura R, Amemiya S, Kanazawa Y. Association of polymorphism in the interferon gamma gene with IDDM. *Diabetologia* 1994; **37:** 1159-62.
19. Jahromi M, Millward A, Demaine A. A CA repeat polymorphism of the IFN-gamma gene is associated with susceptibility to type 1 diabetes. *J Interferon Cytokine Res* 2000; **20:** 187-90.
20. Khani-Hanjani A, Lacaille D, Hoar D*, et al.* Association between dinucleotide repeat in non-coding region of interferon-gamma gene and susceptibility to, and severity of, rheumatoid arthritis. *Lancet* 2000; **356:** 820-5.
21. Vandenbroeck K, Opdenakker G, Goris A, Murru R, Billiau A, Marrosu MG. Interferon-gamma gene polymorphism-associated risk for multiple sclerosis in Sardinia. *Ann Neurol* 1998; **44:** 841-2.
22. Goris A, Epplen C, Fiten P*, et al.* Analysis of an IFN-gamma gene (IFNG) polymorphism in multiple sclerosis in Europe: effect of population structure on association with disease. *J Interferon Cytokine Res* 1999; **19:** 1037-46.
23. van Belzen MJ, Meijer JW, Sandkuijl LA*, et al.* A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* 2003; **125:** 1032-41.
24. Greco L, Corazza G, Babron MC*, et al.* Genome search in celiac disease. *Am J Hum Genet* 1998; **62:** 669-75.

25. King AL, Yiannakou JY, Brett PM*, et al.* A genome-wide family-based linkage study of coeliac disease. *Ann Hum Genet* 2000; **64:** 479-90.
26. Naluai AT, Nilsson S, Gudjonsdottir AH*, et al.* 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* 2001; **9:** 938-44.
27. Neuhausen SL, Feolo M, Camp NJ, Farnham J, Book L, Zone JJ. Genome-wide linkage analysis for celiac disease in North American families. *Am J Med Genet* 2002; **111:** 1-9.
28. Popat S, Bevan S, Braegger CP*, et al.* Genome screening of coeliac disease. *J Med Genet* 2002; **39:** 328-31.
29. Woolley N, Holopainen P, Ollikainen V*, et al.* A new locus for coeliac disease mapped to chromosome 15 in a population isolate. *Hum Genet* 2002; **111:** 40-5.
30. Zhong F, McCombs CC, Olson JM*, et al.* An autosomal screen for genes that predispose to celiac disease in the western counties of Ireland. *Nat Genet* 1996; **14:** 329-33.
31. Liu J, Juo SH, Holopainen P*, et al.* Genomewide linkage analysis of celiac disease in Finnish families. *Am J Hum Genet* 2002; **70:** 51-9.

Chapter 8

The HLA region and coeliac disease

Defining the contribution of the HLA region to DQ2-positive coeliac disease patients

M.J. van Belzen, B.P.C. Koeleman, J.B.A. Crusius, H. Chon, J.W.R. Meijer, A.F.J. Bardoel, P.L. Pearson, L.A. Sandkuijl, R.H.J. Houwen and C. Wijmenga

Submitted for publication.

Summary

The major genetic susceptibility to coeliac disease is contributed by the HLA region. The primary association is with the HLA-DQ2 molecule, encoded by the DQA1*05 and DQB1*02 alleles, which is expressed in over 90% of patients. The aim of our study was to perform an extensive scan of the entire HLA region to determine whether there is any evidence for the presence of additional HLA susceptibility genes, acting independently of DQ2. Sixteen microsatellite markers and the DQA1 and DQB1 genes were genotyped in 120 DQ2-positive simplex coeliac disease families and in 86 DQ2-positive simplex control families. Allele frequencies of markers on phase-known DQ2-positive haplotypes transmitted to patients were compared to a combined group of DQ2-positive non-transmitted and control haplotypes, thereby controlling for the DQ2 contribution. No significant differences at any of the marker loci were detected, suggesting that DQ2 is the major HLA susceptibility locus for coeliac disease on DQ2-haplotypes. Individuals homozygous for DQ2 or heterozygous for DQA1*05-DQB1*02/DQA1*0201-DQB1*02 were found to be at five-fold increased risk for development of coeliac disease ($P < 10^{-8}$). This risk was conferred by the presence of a second DQB1*02 allele next to one DQA1*05-DQB1*02 haplotype, independently of the second DQA1 allele.

Introduction

Coeliac disease is a common food intolerance in humans, with a prevalence estimated to be as high as 0.5%-0.3% in the Netherlands.[1,2] The disease is characterized by lesions of the small intestine with partial to total villous atrophy, crypt hyperplasia and invasion of lymphocytes into the gut epithelium and lamina propria.[3,4] The main clinical symptoms include chronic diarrhoea and growth retardation, but abdominal pain, anaemia, osteopenia, and chronic fatigue may also occur.[5] However, most patients show only some of these symptoms, while others are monosymptomatic or have no symptoms at all. Coeliac disease is caused by dietary intake of gluten peptides from wheat and related proteins from barley and rye. A gluten-free diet usually results in recovery of the small intestinal lesions and disappearance of the clinical symptoms.

Coeliac disease is strongly associated to the human leukocyte antigen (HLA) region. It has been well established that the primary association is with HLA-DQ2, with over 90% of patients expressing this molecule.[6] There is substantial evidence for

involvement of DQ2 in coeliac disease pathogenesis. Gluten-derived peptides are modified by the enzyme tissue transglutaminase, which improves binding to DQ2 on the surface of antigen presenting cells. These complexes are recognized by gluten-specific T cells isolated from small intestinal tissue of coeliac disease patients.[7,8] Most of the DQ2-negative patients express the HLA-DQ8 molecule, which is also capable of binding gluten-derived peptides with subsequent activation of gluten-specific T cells.[7]

The heterodimeric DQ2 protein is encoded by the HLA-DQA1*05 and HLA-DQB1*02 alleles, in either the *cis* or the *trans* configuration. In North European populations, the DQA1*0501 and DQB1*02 alleles are frequently present on the extended HLA-B8-DR3-DQ2 haplotype.[6,9] This haplotype has also been shown to be associated with other autoimmune disorders, including type 1 diabetes mellitus, systemic lupus erythematosus, Graves' disease, Hashimoto's disease, and myasthenia gravis, suggesting that the genes on this haplotype are involved in autoimmunity in general (for a review, see[10]). In coeliac disease it was shown that different DQ2 genotypes account for different disease risks. In particular, the homozygous DQA1*05-DQB1*02/DQA1*05-DQB1*02 (DR3/3) and the heterozygous DQA1*05-DQB1*02/DQA1*0201-DQB1*02 (DR3/7) genotypes were shown to be associated with increased risk.[11-14]

The extended HLA-DR3-DQ2 haplotype includes many other genes that play a role in the immune response and it cannot be excluded that another HLA gene also confers increased risk to coeliac disease. The HLA region is known to display extensive linkage disequilibrium (LD). Therefore, one may expect that specific alleles at various loci in this region will show an increased frequency in coeliac disease patients, not because those alleles enhance or complement the DQ2 risk, but simply because they are in LD with DQ2. However, it has been suggested that non-class II loci also predispose to coeliac disease, independently of DQ2.

The aim of this study was to test whether there was evidence for the presence of additional HLA susceptibility loci for coeliac disease in DQ2-positive patients of Dutch origin. Sixteen markers, covering the entire HLA region and flanking regions, were genotyped in simplex coeliac disease and in control families. In this way, phase-known DQ2-positive haplotypes from cases and controls could be generated and tested for association. In addition, the effect of different DQ2 genotypes to coeliac disease risk was evaluated.

Subjects and methods

Subjects

The case families consisted of 120 unrelated DQ2-positive coeliac disease patients with both parents available. The patients had a mean age of 17 years and 65% were female. The diagnosis of all patients was confirmed by histological re-evaluation of the initial small-intestinal biopsy specimens (JWRM). All patients presented with partial to total villous atrophy in the presence of intraepithelial lymphocytosis and crypt hyperplasia. Control DQ2 haplotypes were derived from 86 control families without a history of coeliac disease. The control families were selected for the presence of at least one parent carrying DQ2 in *cis* and also consisted of one child and both parents. The DQ2-positive controls had a mean age of 48 years and 50% were female. Individuals from the case and control families were all Caucasians of Dutch origin. The study was approved by the Medical Ethics Committee of the University Medical Centre Utrecht and written informed consent was obtained from all the participants.

Table 1. Position, order and primers of the marker loci.

Marker	Position (Mb)[a]	Forward primer (5'-3')	Reverse primer (5'-3')
D6S291	36.26	CTCAGAGGATGCCATGTCTAAAATA	GGGGATGACGAATTATTCACTAACT
D6S2414	32.85	AACTGGGCTGAGATGTACCA	AAGGAGAGGAATGTGTGTGC
TAP1	32.82	GCTTTGATCTCCCCCCTC	GGACAATATTTTGCTCCTGAGG
STR2	32.71	GATTCATGAGCCAAGAACCC	ATAATGCCATTCAATGTAAGC
DQB1	32.63	ATGTGCTACTTCACCAACGG	CTGGTAGTTGTGTCTGCACAC
DQA1	32.61	ATGGTGTAAACTTGTACCAGT	TTGGTAGCAGCGGTAGAGTTG
D6S273	31.74	GCAACTTTTCTGTCAATCCA	ACCAAACTTCAAATTTTCGG
TNFa	31.59	CCTCTAGATTTCATCCAGCCACAG	CCTCCCCCTCTCTCCCCTGC
MICB	31.53	AGTGTTTTCCATTGCAGGCG	ATGGGCAAGACTTCAATGGC
MICA	31.43	CCTTTTTTTCAGGGAAAGTGC	CTTACCATCTCCAGAAACTGC
MIB	31.4	ACCACAGTCTCTATCAGTCC	TCTACCATGACCCCCTTCC
STR1	31.31	AGCATATCTGCCATTTGGCC	GAAACTTGGGCAATGAGTCC
D6S2700	30.84	AAAAGGAGGAAGAGCCACGGAG	CTGTGAGTAGTAAGAACCCCC
D6S2704	30.23	GAGCACAATATCTGGTCTGCTGC	TTTTGCCACTCTGGAGGATGG
D6S2707	29.77	CAGTTTCGCAACCTGTTTGCC	TCTGATAAGAGATTAATATCCAG
D6S105	27.83	GCCCTATAAAATCCTAATTAAC	GAAGGAGAATTGTAATTCCG
D6S2223	27.72	AATGTTAAGTAACAAACTAGAGTAC	ACTCCAGCCTGGGCAATAGAGC
D6S1281	25.35	GATGCCACGTTTTAAAATGC	AGAAGCAGCTGTGCTTTGTT

[a] The position of the markers was obtained by blasting the primer sequences against the Ensembl database, June 2003 release.

Genotyping the HLA loci

A total of 18 loci were genotyped in the case and control families (Table 1 and Figure 1). The DQA1 and DQB1 genes were typed as described before, and primer sequences are listed in Table 1.[15] Fourteen microsatellite markers, spanning the entire HLA region, and two flanking markers (D6S291 and D6S1281) were amplified in multiplex PCR reactions (Table 1). The reaction volume of 10 µl contained 25 ng of DNA, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 50 ng fluorescence-labelled primer and 0.4 U AmpiTaq Gold (PE Applied Biosystems, Foster City). The PCR products were pooled and separated on a 3700 DNA sequencer and analysed by Genescan 3.5 and Genotyper 2.0 software (all from PE Applied Biosystems). All genotypes were checked independently by two researchers (MJvB and AFJB). Primers sequences were obtained from the Genome Database (www.gdb.org), except for markers STR1, STR2, MICB, MIB, TNFa and TAP1, which were designed in our lab. Primer sequences for all loci were blasted against the Ensembl database (June 2003 release) to determine the correct order of the loci.

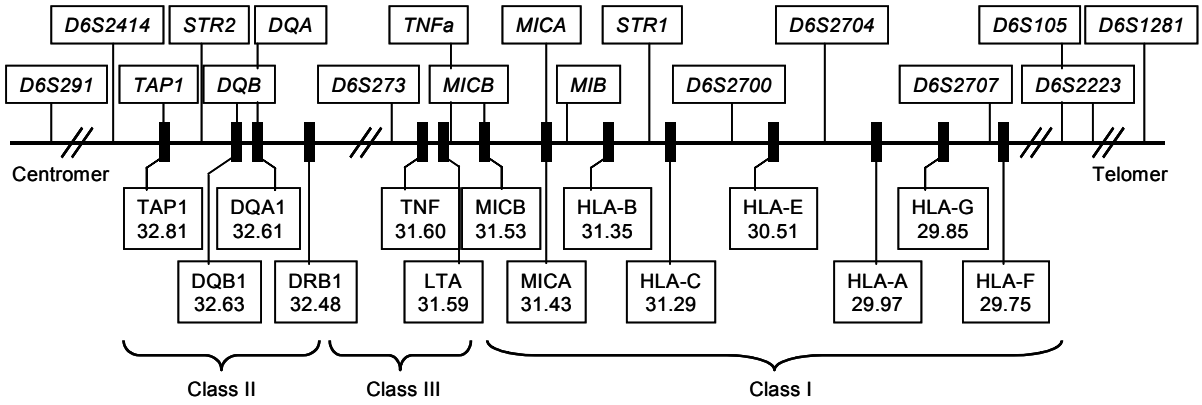


Figure 1. Overview of major genes in the HLA region and marker loci from this study. Vertical bars indicate genes, with names and locations depicted in the lower half of the figure. The location of the genes was based on the June 2003 release of the Ensembl human genome map. Marker loci are depicted in *italics* in the upper half of the figure.

Statistical analysis

Phase-known haplotypes from the case and control families were constructed with the SHOWHAPLO program (F. Dudbridge). Four patients turned out to be DQ2-positive, but in the *trans* configuration and they were therefore excluded from the analysis. Six of

the 116 *cis* DQ2-positive patients were not informative at either the DQA1 or DQB1 locus, resulting in 110 informative case families that were eventually included in the analysis. These 110 cases carried in total 150 DQ2-positive haplotypes (T haplotypes), while 36 DQ2-positive haplotypes were not transmitted to the cases (NT haplotypes). The control families contained 93 *cis* DQ2-positive parents who carried together 103 DQ2-positive haplotypes (control haplotypes).

Allele frequencies at the 16 microsatellite marker loci were determined on the T, NT and control haplotypes. The overall allele distribution on T haplotypes was compared to the combined group of NT and control haplotypes. In addition, alleles with a frequency ≥10% were tested separately for association. Statistical significance was determined by χ^2 analysis. TDT analysis, conditioned on the presence of DQ2 was performed in the case families by the CETDT program.[16]

The effect of DQ2 genotype to coeliac disease risk was determined in cases and controls. Association of the different DQ2 genotypes was tested for significance by χ^2 analysis with 1 df, by testing each DQ2 genotype against all other DQ2 genotypes. Odds ratios were calculated using Woolf's method with Haldane's correction relative to DQA1*05-DQB1*02/DQA1*X-DQB1*X (DR3/X) as reference genotype.

Results

Comparison of microsatellite marker loci on cis DQ2-positive haplotypes

Allele frequencies of all 16 microsatellite markers were determined on the T, NT and control haplotypes. The overall allele distribution was not significantly different between T and the combined group of NT and control haplotypes at any of the marker loci (data not shown). The most frequent allele on the T haplotypes was determined and the frequency of this allele was also compared to the combined group of NT and control haplotypes. These alleles are markers of the B8-DR3-DQ2 haplotype, characterized by alleles TNFa*2, MICB*10, MICA*3 and MIB*11 (Table 2).[17,18] This haplotype is the major DQ2 haplotype in patients and controls, with a frequency >75%. No evidence for significant association of these B8-DR3-DQ2 specific alleles was found. Other alleles with a frequency of 10% or more were also tested for association with coeliac disease (Table 3). These less frequent alleles could be compared for eight markers, but no significant evidence for association with coeliac disease was found. Only one marginally

significant result was obtained for allele 4 of marker D6S2707, but the significance was lost after correction for multiple testing.

Table 2. Comparison of the most frequent allele of each microsatellite marker on DQ2-positive haplotypes transmitted (T) and non-transmitted (NT) to patients and on control haplotypes.

Locus	Allele (bp)	T (N = 150)	NT (N = 36)	Controls (N = 103)	NT + Controls (N = 139)	P-value T vs NT + Controls
D6S291	2 (201)	57	13	44	57	0.60
D6S2414	2 (170)	93	24	53	77	0.19
TAP1	3 (192)	106	26	66	92	0.27
STR2	6 (108)	127	28	81	109	0.19
D6S273	7 (140)	110	22	82	104	0.62
TNFa	2 (104)	121	24	86	110	0.67
MICB	10 (292)	114	24	77	101	0.68
MICA	3 (187)	117	26	79	105	0.83
MIB	11 (354)	106	26	78	104	0.64
STR1	6 (127)	110	26	76	102	0.73
D6S2700	3 (227)	109	30	72	102	0.36
D6S2704	7 (165)	75	17	56	73	0.73
D6S2707	9 (312)	83	20	56	76	0.78
D6S105	5 (119)	65	13	44	57	0.70
D6S2223	3 (170)	110	25	75	100	0.75
D6S1281	4 (194)	39	7	27	34	0.67

Table 3. Comparison of other alleles with frequencies ≥10% on DQ2-positive haplotypes in cases (T) and a combined group of non-transmitted and control haplotypes (NT + controls).

Locus	Allele (bp)	T (N = 150)	NT + Controls (N = 139)	P-value T vs NT + Controls
D6S291	1 (199)	38	34	0.86
	6 (209)	15	16	0.68
D6S2414	3 (174)	34	39	0.30
TAP1	2 (190)	28	34	0.15
MICB	1 (272)	15	11	0.57
D6S2704	8 (167)	23	17	0.42
D6S2707	4 (302)	18	6	0.023[a]
D6S105	6 (121)	15	8	0.18
	7 (123)	21	28	0.16
	8 (125)	18	10	0.17
D6S1281	5 (198)	37	27	0.23
	6 (202)	30	35	0.33

[a] This P-value is not corrected for multiple testing.

Conditional extended TDT (CETDT)

CETDT analysis in the case families, conditioned on DQ2, did not detect any alleles at the marker loci that showed significant distortion of random transmission.

Effect of different DQ2 genotypes

The DQ2 genotypes carried by the 110 DQ2-positive cases and 93 DQ2-positive controls are shown in Table 4. The homozygous DQA1*05-DQB1*02/DQA1*05-DQB1*02 (DR3/3) and heterozygous DQA1*05-DQB1*02/DQA1*0201-DQB1*02 (DR3/7) genotypes both conferred a five-fold increased risk for coeliac disease. Homozygosity for DQB1*02 was strongly associated with coeliac disease ($P < 10^{-8}$).

Table 4. DQ2 genotype frequencies in *cis* DQ2-positive coeliac disease patients and controls.

DQ2 genotype[a]	DR type[b]	Cases (N = 110)	Controls (N = 93)	P-value[c]	OR (95% CI)[d]
DQA1*05-DQB1*02/ DQA1*05-DQB1*02	DR3/3	40 (36%)	10 (11%)	0.00002	5.54 (2.5-12.1)
DQA1*05-DQB1*02/ DQA1*0201-DQB1*02	DR3/7	24 (22%)	6 (6%)	0.002	5.31 (2.1-13.5)
DQA1*05-DQB1*02/ DQA1*0301- DQB1*0302	DR3/4	7 (6%)	13 (14%)		
DQA1*05-DQB1*02/ DQA1*05-DQB1*0301	DR3/5	3 (3%)	10 (11%)		
DQA1*05-DQB1*02/ DQA1*X-DQB1*X	DR3/X	36 (33%)	54 (58%)		

[a] Haplotypes are phase-known as parents or children were also genotyped. DQA1*X-DQB1*X refers to anything except those haplotypes listed in Table 4.

[b] DR genotype was not acquired, but derived from alleles at the DQA1 and DQB1 locus. X refers to anything except DR3, 4, 5 or 7.

[c] P-values for association with coeliac disease were calculated by testing each DQ2 genotype against all other DQ2 genotypes. P-value for the presence of two DQB1*02 alleles $<10^{-8}$.

[d] Odds ratios were calculated relative to the DR3/X type as reference. OR for the presence of two DQB1*02 alleles = 5.68 (95% CI 2.9-11.2).

Discussion

We investigated whether there was support for the presence of additional risk factors for coeliac disease in the HLA region, independent of DQ2. As the HLA region exhibits strong LD, several approaches have been suggested to correct for this. The homozygous

parent TDT was first applied to search for HLA genes in type 1 diabetes mellitus.[19] Only transmissions from parents homozygous for DQ2 and DR3 and heterozygous at the test loci were included in the analysis. This approach was later extended to a case-control design, in which only DR3 homozygous cases and controls were included.[20] Although both methods control elegantly for the existing LD, they have very little power. Only individuals homozygous for DQ2 or DR3 are informative, which leads to exclusion of the majority of the data set. Recently, a TDT approach using affected family-based controls (AFBAC) was applied to two large data sets of coeliac disease families.[21,22] This approach allows for construction of phase-known haplotypes, in which transmitted DQ2-positive haplotypes are compared to non-transmitted DQ2-positive haplotypes. A major advantage is that all DQ2-positive haplotypes are included in the analysis. However, less then 20% of DQ2-positive haplotypes were non-transmitted, so it requires a rather large data set to provide sufficient power.[21,22]

In view of this all, we chose to use a combined case-control and AFBAC approach, as well as CETDT, to maximize the power of our data set, which is of moderate size. By also genotyping the parents of cases and children of controls, we were able to obtain phase-known haplotypes in both groups. Hence, all the DQ2-positive haplotypes could be used in the case-control analysis. The AFBAC DQ2-positive haplotypes were combined with the DQ2-positive control haplotypes into one control group since both should represent population control haplotypes. This case-control design greatly increased the power of our study, as an AFBAC case-control approach would have resulted in only 36 DQ2-positive control haplotypes, compared to 139 when using the combined group of AFBAC and control haplotypes.

We performed an extensive scan of the HLA region using microsatellite markers, but no evidence for independent association between any of the loci and coeliac disease was found. Likewise, no significant differences were present when comparing T and NT haplotypes, or T and control haplotypes separately (data not shown). Three other extensive screens of the HLA region in search for additional risk loci have been performed in coeliac disease. Two of them did not find significant association, independently of DQ2, at any of the microsatellite loci either.[21,23] The presence of an additional risk locus in the *MIC* gene region was suggested by the third study.[22] Association of the *MICA* gene in a DQ2-positive population has been reported twice,

but the control DQ2 groups were small in both studies.[24,25] In addition, independent association of a single nucleotide polymorphism (SNP) in the *TNF* gene region has been reported by several groups.[21,26-28] These results indicate the possibility of additional HLA risk loci in these regions, but large collections of phase-known DQ2-positive patients and controls are necessary to provide unambiguous evidence. Furthermore, there is a possibility that a SNP conferring increased disease risk may not be detected by association of nearby microsatellite markers because of their high mutation rate. Therefore, SNPs may be the preferred type of polymorphism for studying the presence of additional HLA susceptibility loci.

The presence of an independent, additional risk locus for coeliac disease, located telomeric to the HLA class I region, has been suggested in a case-control study using DR3 homozygous patients and controls.[20] Allele 3 of marker D6S2223 was significantly less frequent in cases. However, we were unable to confirm these findings in our data set. Furthermore, we found no evidence for association of allele 3 in an unstratified analysis of both case-control (76% vs 74%, P = NS) and TDT (43 T vs 39 NT, P = NS) data (data not shown). Three other studies, all using large data sets, were also unable to confirm association of D6S2223 with coeliac disease.[21-23] These results indicate that it is unlikely that a gene near D6S2223 predisposes to coeliac disease. The previously reported association may have occurred by chance due to the rather small sample size of 46 patients.

The majority of the DQ2-positive haplotypes in cases and controls consisted of B8-DR3-DQ2 haplotypes (see also Table 2).[17,18] We were therefore unable to determine whether there was a locus specifically associated with coeliac disease on other backgrounds. For example, the B18-DR3-DQ2 extended haplotype, characterized by alleles TNFa*1, MICB*1, MICA*1 and MIB*1, was rare in our cohorts with 9 T, 1 NT and 5 control haplotypes (data not shown).[17,18] This haplotype is carried by 84% of DQ2-positive Sardinian coeliac disease patients, and this population is therefore more suitable for studying the presence of additional HLA risk loci on B18-DR3-DQ2 haplotyes.[29]

In addition to our search for the presence of additional non-class II HLA loci, our data set also enabled us to establish the risk of different DQ2 genotypes to coeliac disease in the Dutch population. Homozygous DQA1*05-DQB1*02/DQA1*05-DQB1*02 and heterozygous DQA1*05-DQB1*02/DQA1*0201-DQB1*02 individuals were at five-fold

increased risk. This risk is attributed by the presence of a second DQB1*02 allele and seems independent of the second DQA1 allele, since the odds ratio for homozygosity of DQB1*02 is almost equal to those for both risk genotypes. Similar risks were observed in other populations, although these studies were conducted in an ordinary case-control setting, in which haplotypes had to be estimated.[12-14] Recently, a family study using a TDT approach with phase-known haplotypes also demonstrated increased risk for these two genotypes.[11] A possible explanation for the increased risk may reside in number of DQ molecules capable of gluten presentation that arise from each genotype. The homozygous DQA1*05-DQB1*02 genotype produces 100% DQ2 molecules. The heterozygous DQA1*05 DQB1*02/DQA1*0201-DQB1*02 genotype produces only 50% DQ2 molecules and the other 50% is comprised by the α1*0201-β1*02 heterodimer. However, the α1*0201-β1*02 molecule was shown to be able to present certain gluten epitopes to T cells as well, thereby implicating this molecule in the pathogenesis of coeliac disease (Vader et al., manuscript submitted). The DQA1*05 DQB1*02/DQA1*X-DQB1*X genotype results in just 25% DQ2 molecules, which may account for the lower disease risk of this genotype.

In conclusion, in this study in coeliac disease using phase-known DQ2-positive case and control haplotypes, we were not able to find support for the presence of an additional HLA susceptibility gene, acting independently of DQ2. We were able to confirm the increased risk conferred by homozygosity for the DQB1*02 allele, which is most likely due to a combination of the number of DQ2 molecules expressed and the gluten presenting capacity of the heterodimer encoded by DQA1*0201 and DQB1*02.

Acknowledgements

We are grateful to Jessica Bader for DQA1 and DQB1 genotyping of the case and control families. We thank Jackie Senior for improving the manuscript. The study was financially supported by a grant from the Dutch Digestive Diseases Foundation (WS 97-44).

This paper is dedicated to the memory of Lodewijk Sandkuijl who unexpectedly died at December 4, 2002. He participated in this project from the beginning and his

contribution to the design of the study and analysis of the data was of great importance to us.

References

1. Csizmadia CG, Mearin ML, von Blomberg BM, Brand R, Verloove-Vanhorick SP. An iceberg of childhood coeliac disease in the Netherlands. *Lancet* 1999; **353:** 813-4.
2. Rostami K, Mulder CJ, Werre JM*, et al.* High prevalence of celiac disease in apparently healthy blood donors suggests a high prevalence of undiagnosed celiac disease in the Dutch population. *Scand J Gastroenterol* 1999; **34:** 276-9.
3. Rostami K, Kerckhaert J, Tiemessen R, von Blomberg BM, Meijer JW, Mulder CJ. Sensitivity of antiendomysium and antigliadin antibodies in untreated celiac disease: disappointing in clinical practice. *Am J Gastroenterol* 1999; **94:** 888-94.
4. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 1992; **102:** 330-54.
5. Farrell RJ, Kelly CP. Celiac sprue. *N Engl J Med* 2002; **346:** 180-8.
6. Sollid LM, Thorsby E. HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 1993; **105:** 910-22.
7. van de Wal Y, Kooy Y, van Veelen P*, et al.* Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. *J Immunol* 1998; **161:** 1585-8.
8. Molberg O, McAdam SN, Korner R*, et al.* Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat Med* 1998; **4:** 713-7.
9. Alper CA, Fleischnick E, Awdeh Z, Katz AJ, Yunis EJ. Extended major histocompatibility complex haplotypes in patients with gluten-sensitive enteropathy. *J Clin Invest* 1987; **79:** 251-6.
10. Candore G, Lio D, Colonna Romano G, Caruso C. Pathogenesis of autoimmune diseases associated with 8.1 ancestral haplotype: effect of multiple gene interactions. *Autoimmun Rev* 2002; **1:** 29-35.
11. Louka AS, Nilsson S, Olsson M*, et al.* HLA in coeliac disease families: a novel test of risk modification by the 'other' haplotype when at least one DQA1*05-DQB1*02 haplotype is carried. *Tissue Antigens* 2002; **60:** 147-54.
12. Ploski R, Ek J, Thorsby E, Sollid LM. 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* 1993; **41:** 173-7.
13. Mearin ML, Biemond I, Pena AS*, et al.* HLA-DR phenotypes in Spanish coeliac children: their contribution to the understanding of the genetics of the disease. *Gut* 1983; **24:** 532-7.
14. Mearin ML, Bouquet J, Mourad N*, et al.* HLA-DR antigens and phenotypes in Dutch coeliac children and their families. *Clin Genet* 1985; **27:** 45-50.
15. Carrington M, Miller T, White M*, et al.* Typing of HLA-DQA1 and DQB1 using DNA single-strand conformation polymorphism. *Hum Immunol* 1992; **33:** 208-12.
16. Koeleman BP, Dudbridge F, Cordell HJ, Todd JA. Adaptation of the extended transmission/disequilibrium test to distinguish disease associations of multiple loci: the Conditional Extended Transmission/Disequilibrium Test. *Ann Hum Genet* 2000; **64:** 207-13.

17. Bolognesi E, Dalfonso S, Rolando V, Fasano ME, Pratico L, Momigliano-Richiardi P. MICA and MICB microsatellite alleles in HLA extended haplotypes. *Eur J Immunogenet* 2001; **28:** 523-30.
18. Grimaldi MC, Clayton J, Pontarotti P, Cambon-Thomsen A, Crouau-Roy B. New highly polymorphic microsatellite marker in linkage disequilibrium with HLA-B. *Hum Immunol* 1996; **51:** 89-94.
19. Lie BA, Todd JA, Pociot F*, et al.* The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene. *Am J Hum Genet* 1999; **64:** 793-800.
20. Lie BA, Sollid LM, Ascher H*, et al.* A gene telomeric of the HLA class I region is involved in predisposition to both type 1 diabetes and coeliac disease. *Tissue Antigens* 1999; **54:** 162-8.
21. Louka AS, Lie BA, Talseth B*, et al.* Coeliac disease patients carry conserved HLA-DR3-DQ2 haplotypes revealed by association of TNF alleles. *Immunogenetics* 2003; **55:** 339-43.
22. Bolognesi E, Karell K, Percopo S*, et al.* Additional factor in some HLA DR3/DQ2 haplotypes confers a fourfold increased genetic risk of celiac disease. *Tissue Antigens* 2003; **61:** 308-16.
23. Louka AS, Moodie SJ, Karell K*, et al.* A collaborative European search for non-DQA1*05-DQB1*02 celiac disease loci on HLA-DR3 haplotypes: analysis of transmission from homozygous parents. *Hum Immunol* 2003; **64:** 350-8.
24. Lopez-Vazquez A, Rodrigo L, Fuentes D*, et al.* MHC class I chain related gene A (MICA) modulates the development of coeliac disease in patients with the high risk heterodimer DQA1*0501/DQB1*0201. *Gut* 2002; **50:** 336-40.
25. Rueda B, Pascual M, Lopez-Nevot MA*, et al.* Association of MICA-A5.1 allele with susceptibility to celiac disease in a family study. *Am J Gastroenterol* 2003; **98:** 359-62.
26. McManus R, Wilson AG, Mansfield J, Weir DG, Duff GW, Kelleher D. TNF2, a polymorphism of the tumour necrosis-alpha gene promoter, is a component of the celiac disease major histocompatibility complex haplotype. *Eur J Immunol* 1996; **26:** 2113-8.
27. de la Concha EG, Fernandez-Arquero M, Vigil P*, et al.* Celiac disease and TNF promoter polymorphisms. *Hum Immunol* 2000; **61:** 513-7.
28. Garrote JA, Arranz E, Telleria JJ, Castro J, Calvo C, Blanco-Quiros A. TNF alpha and LT alpha gene polymorphisms as additional markers of celiac disease susceptibility in a DQ2-positive population. *Immunogenetics* 2002; **54:** 551-5.
29. Congia M, Frau F, Lampis R*, et al.* A high frequency of the A30, B18, DR3, DRw52, DQw2 extended haplotype in Sardinian celiac disease patients: further evidence that disease susceptibility is conferred by DQ A1*0501, B1*0201. *Tissue Antigens* 1992; **39:** 78-83.

Chapter 9

General Discussion

Coeliac disease: a common but underrecognized genetic disorder

The prevalence of coeliac disease in the Netherlands has been estimated by two large population screens at about 1 in 250 (0.4%).[1,2] Consequently, we would expect approximately 64,000 coeliac disease patients in the Netherlands, based on 16,000,000 inhabitants. However, at the beginning of 2003, the Dutch Coeliac Disease Foundation had only 6000 members.[3] Most patients join the foundation for support and dietary advice, so their membership is a good indication of the total number of patients diagnosed with coeliac disease. The large discrepancy indicates that the vast majority of patients (90%) remain undiagnosed. The undiagnosed group includes patients that are asymptomatic and, hence, will not be tested for coeliac disease unless perhaps when other family members have been diagnosed with the disease. However, many patients suffering from non-specific complaints like chronic fatigue, anaemia or abdominal pain remain unrecognised. Clinicians, not only specialists but also general practitioners, should always be aware of the possibility of coeliac disease in patients with vague complaints that could be associated with the disease (see also Chapter 1).

Coeliac disease is a disorder with a strong genetic aetiology. The sibling recurrence risk is about 10%,[4-12] which leads to a sibling relative risk (λ_s) of 25 for the Dutch population.[13,14] This λ_s value is relatively high and therefore provides a good opportunity for the identification at least some of the coeliac disease susceptibility genes in a data set of average size. The identification of these genes will be of great importance for understanding the pathogenesis of the disease. It will provide new insights into the pathways that are impaired in the disease, which may assist diagnosis of the disease and even lead to new therapeutic targets. The aim of this PhD project was to investigate the genetic factors that underlie coeliac disease in the Dutch population. For this purpose, three different cohorts of coeliac disease patients were created: a cohort of affected sibpairs to localise the disease genes (Chapter 2), and a case-control cohort and a cohort of parent-case trios to perform association analysis of candidate regions (Chapter 4), and candidate genes (Chapters 5-7) and to further establish the contribution of the HLA region (Chapter 8). Great effort was put into ensuring homogeneity of the diagnoses of the patients included in all three cohorts. The initial biopsy specimens of all patients were re-evaluated by Dr. J.W.R. Meijer (an experienced pathologist of the Rijnstate hospital,

Arnhem). All three cohorts consisted of patients with a Marsh III lesion; patients with less severe lesions or whose biopsy specimens were unavailable were not included in these cohorts.

The initial goal was to collect 100 affected sibpairs of Dutch origin and perform the first genome-wide screen in Dutch coeliac disease patients. At the beginning of the project in 1998, about 3000 patients were diagnosed with coeliac disease. Based on a sibling recurrence risk of 10%, there were expected to be approximately 300 families with two affected children in the Netherlands. But after five years of extensive, nationwide searching, only 82 families with 101 affected sibpairs have been collected while at the same time the total number of diagnosed patients has doubled. Since the patients and their families were very cooperative, the apparent shortage of families was probably due to an overestimation of their occurrence. Recurrence risks for first-degree relatives were based on family screening studies.[4,6-12] The actual numbers of diagnosed siblings is lower since the disease also remains unrecognised in family members. Furthermore, several families had to be excluded because some patients presented with the less severe Marsh I and II lesions, some patients had been diagnosed many years ago and their biopsy specimens were no longer available, and some patients had not been diagnosed by a small intestinal biopsy. Taking all these considerations into account, our affected sibpair cohort probably contains the majority of families with multiple affected Marsh III siblings present in the Netherlands. The case-control cohort (216 patients and matched controls) and parent-case trios (122 families) were fairly easy to obtain via the Paediatric Gastroenterology department of our own hospital (Dr. R.H.J. Houwen, UMC Utrecht) and via the Gastroenterology department of the Rijnstate hospital in Arnhem (Prof. dr. C.J.J. Mulder, currently at the VUMC, Amsterdam). Both cohorts are of moderate size as highest priority was given to collecting affected sibpairs, but they should be large enough to detect association of genes conferring moderate to large risks (relative risk (RR) ≥2). However, it should be relatively easy to extend these cohorts for future experiments so that genes with small relative risks can be detected as well.

Coeliac disease and the HLA region

Coeliac disease is strongly associated with the HLA region

In the early 1970's it was already noticed that coeliac disease was associated with the human leukocyte antigen (HLA) region. First reported associations were with the B8 antigen from the class I region, but later a stronger association with the class II DR3 antigen was found (Figure 1).[13] Now it is well established that the primary association is with the class II antigen DQ2. Over 90% of patients express this heterodimeric molecule, encoded by the DQA1*05 and DQB1*02 alleles. Most patients of northern European origin carry these alleles on the extended HLA-B8-DR3-DQ2 haplotype, which explains the early reported associations with HLA-B8 and HLA-DR3.[13,15] The occurrence of the extended B8-DR3-DQ2 haplotype in the majority of patients has initiated a search for additional susceptibility genes in the HLA region that increase the disease risk independently of DQ2. This haplotype has certain functional characteristics due to the presence of specific alleles at regulatory loci, including high expression of tumour necrosis factor (TNF) α and the production of a truncated MICA protein.[16,17] Four extensive scans of the HLA region with microsatellite markers were performed to localize these possible additional risk loci. Three of them, including our study in a case-control cohort using phase-known DQ2 positive haplotypes, did not find association of any marker independently of DQ2.[18,19] The presence of an additional risk locus in the MIC gene region was suggested in one study.[20] In addition, independent association of a single nucleotide polymorphism (SNP) in the TNF gene region has been reported by several groups.[19,21-23] There is a possibility that a SNP conferring increased disease risk may not be detected by association with nearby microsatellite markers because of their high mutation rate. Therefore, SNPs may have been a better choice for studying the presence of additional, non-class II HLA susceptibility loci. However, SNPs only contain two alleles and single SNPs therefore provide low information. Furthermore, genotyping only one SNP will be insufficient to localize the putative additional risk locus, as this SNP may be in linkage disequilibrium (LD) with another variant that is actually causing the disease. Therefore, studies should include several polymorphisms in multiple genes and construction of phase-known haplotypes to identify the regions containing additional susceptibility genes. Nevertheless, detection of additional susceptibility loci may be nearly impossible due to extensive LD displayed by the HLA region.

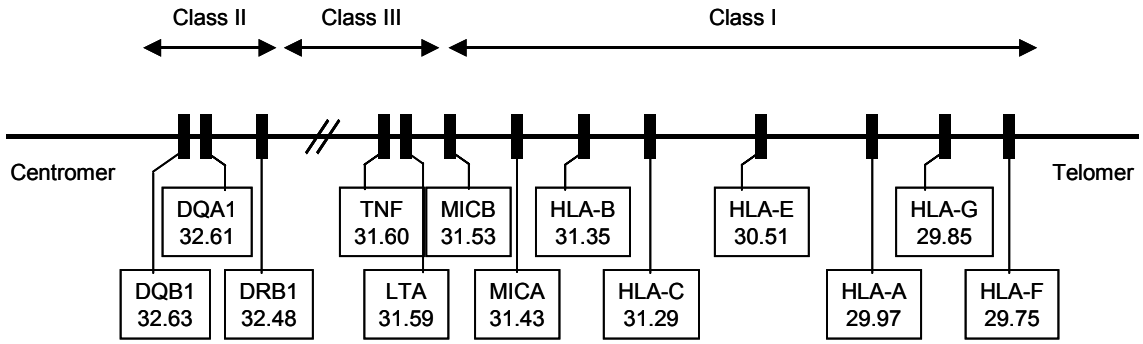


Figure 1. Overview of major genes in the HLA region. Vertical bars indicate genes, with names and locations depicted in the lower half of the figure. The location of the genes was based on the June 2003 release of the Ensembl human genome map.

In patients from certain southern European populations, the DQ2 risk alleles are frequently carried on the extended B18-DR3-DQ2 haplotype.[24] This haplotype shares the DQA1*05, DQB1*02 and DRB1*03 alleles with the B8-DR3-DQ2 haplotype, but differs at most other loci, indicating that the DQ-DR region is essential for development of coeliac disease.[17,25] A third haplotype, the B21-DR3-DQ2 haplotype, was shown to be associated with coeliac disease in Indian patients.[26] The presence of different disease associated DR3-DQ2 haplotypes implies that these haplotypes may harbour unique additional risk loci. To find such loci, association analysis should be focused on different DQ2 subgroups, stratified on the presence of either B8, B18 or B21. Not all populations will be suitable for this type of study, as the distribution of these haplotypes varies in Europe.[13] North European populations can be used for studying the contribution of the B8-DR3 haplotype, while for example the Sardinian population provides a unique possibility to determine the contribution of the B18-DR3 haplotype, as 84% of DQ2-positive patients were shown to carry this haplotype.[24]

The majority of DQ2-negative patients are DQ8-positive and carry the DQA1*0301 and DQB1*0302 alleles on the DR4 haplotype. It would be interesting to determine whether evidence exists for the presence of additional risk loci on this haplotype. However, since DQ2-negative, DQ8-positive patients comprise only a small part of all coeliac disease patients, very large patient groups are required to obtain enough power to reach statistical significance. For example, only two out of 122 (1.6%) patients

in our parent-case trios carried the DQ8 alleles in the absence of DQ2. So, this question can only be addressed by large-scale collaboration between research groups. However, even the large collection of coeliac disease patients by a European consortium contained only 60 DQ2-negative, DQ8-positive patients on a total of 1008 (6%) independent patients. Unfortunately, a scan of the HLA region to search for additional HLA susceptibility loci that act independently of DQ8 has not yet been performed.

*The DQB1*02 gene dosage effect*

Several studies reported that individuals homozygous for DR3/3 and heterozygous for DR3/7 were at higher risk for development of coeliac disease compared to heterozygous DR3/X individuals (where X is not 3 or 7).[27-29] The increased risk was due to the presence of a second DQB1*02 allele, either on a DR3 or DR7 haplotype, next to one DR3 haplotype. This gene dosage effect was confirmed in our cohort, with a five-fold increased risk for both DQA1*05-DQB1*02/DQA1*05-DQB1*02 (DR3/3) homozygous and DQA1*05-DQB1*02/DQA1*0201-DQB1*02 (DR3/7) heterozygous individuals (see Chapter 8).

The increased risk conferred by these two genotypes may reside in the percentage of DQ2 molecules they encode, which is higher than for the DR3/X genotype (Table 1). The similar risks conferred by DR3/3 and DR3/7 are most likely due to functional characteristics of the α1*0201-β1*02 heterodimer, which was recently implicated in gluten presentation to T cells (Vader et al., manuscript submitted). Other genotypes also produce 50% or more DQ2 and DQ8 molecules and may therefore be expected to be associated with an increased disease risk as well (Table 1). No significant evidence in favour of this has been generated, but this may be due to the fact that these genotypes are rare among coeliac disease patients. Furthermore, no data is available on the functional characteristics of the DQ heterodimers they encode. Notably, most of the 60 DQ8 patients studied by the European consortium were either homozygous DQA1*0301-DQB1*0302 (48%) or heterozygous DQA1*0301-DQB1*0302/DQA1*0201-DQB1*02 (18%), but no data on a control DQ8 population was available.[30]

Table 1. Potential DQ alpha-beta chain combinations in coeliac disease associated haplotypes and their gluten presenting capacity.

DQ2 and DQ8 genotypes[a]	DR type	Possible DQ molecules[b]	DQ type[c]	Gluten-presenting capacity	% of gluten-presenting DQ molecules
DQA1*05-DQB1*02 DQA1*05-DQB1*02	3/3	α1*05/β1*02	DQ2	+	100
DQA1*05-DQB1*02/ DQA1*0201-DQB1*02	3/7	α1*05/β1*02	DQ2	+	≥50
		α1*0201/β1*02	–	±[d]	
DQA1*05-DQB1*02 DQA1*X-DQB1*X	3/X	α1*05/β1*02	DQ2	+	25
		α1*05/β1*X	–	–	
		α1*X/β1*02	–	–	
		α1*X/β1*X	–	–	
DQA1*05-DQB1*02 DQA1*05-DQB1*0301	3/5	α1*05/β1*02	DQ2	+	50
		α1*05/β1*0301	–	?	
DQA1*05-DQB1*02 DQA1*0301-DQB1*0302	3/4	α1*05/β1*02	DQ2	+	50
		α1*05/β1*0302	–	?	
		α1*0301/β1*02	–	?	
		α1*0301/β1*0302	DQ8	+	
DQA1*0301-DQB1*0302 DQA1*0301-DQB1*0302	4/4	α1*0301/β1*02	DQ8	+	100
DQA1*0301-DQB1*0302 DQA1*0201-DQB1*02	4/7	α1*0301/β1*0302	DQ8	+	≥25
		α1*0301/β1*02	–	?	
		α1*0201/β1*0302	–	–	
		α1*0201/β1*02	–	±[d]	
DQA1*05-DQB1*0301 DQA1*0201-DQB1*02	5/7	α1*05/β1*0301	–	?	≥25
		α1*05/β1*02	DQ2	+	
		α1*0201/β1*0301	–	–	
		α1*0201/β1*02	–	±[d]	

[a] DQA1*X is not 05 or 0201 and DQB1*X is not 02.

[b] α1*X is not 05 or 0201 and β1*X is not 02.

[c] "–" refers to not DQ2 or DQ8.

[d] This molecule was able to present certain gluten-peptides to T cells, but not all (Vader et al., manuscript submitted).

Are DQ2 and DQ8 essential?

The presence of either DQ2 or DQ8 appears to be necessary for development of coeliac disease. Almost all patients carry these alleles and functional involvement of both DQ molecules in presentation of gluten to the immune system has been demonstrated

unambiguously. Nevertheless, patients sensitive to gluten but negative for DQ2 and DQ8 have been described.[30-32] Of the 1008 coeliac disease patients studied by the European consortium, 61 (6%) were found to be DQ2 and DQ8 negative.[30] However, 57 (93%) of them carried half of the DQ2 heterodimer: 41 (72%) were positive for DQA1*0201-DQB1*02 and 16 (28%) were positive for DQA1*05-DQB1*0301. As mentioned earlier, the molecule encoded by the DQA1*0201 and DQB1*02 alleles has been implicated in gluten presentation to T cells. However, no data has been generated yet on the T cell stimulation capacity of the α1*05-β1*0301 heterodimer. The current data are of great importance for the clinical practice of diagnosing coeliac disease. Negativity for DQ2 and DQ8 should not be used to reject a possible diagnosis of coeliac disease and a small intestinal biopsy should be performed in all individuals who carry at least half of the DQ2 heterodimer and suffer from any complaint compatible with coeliac disease.

Of course, one has to bear in mind that some of the reported DQ2 and DQ8 negative patients may have been incorrectly diagnosed with coeliac disease. After a biopsy specimen is taken, it should be handled with care and oriented properly to avoid artificial damage and misdiagnosis.[33,34] Patients diagnosed with coeliac disease but negative for DQ8, DQA1*05 and DQB1*02 should be subjected to a more extensive diagnostic procedure to confirm the diagnosis. The original criteria, formulated by the European Society of Paediatric Gastroenterology and Nutrition (ESPGAN) during the Interlaken meeting in 1969, may provide a good protocol for these cases.[35] This protocol involves small intestinal biopsies at three different stages. The diagnosis is confirmed upon the presence of villous atrophy on a gluten-containing diet in the first biopsy, recovery of the villous structure on a gluten-free diet, and reoccurrence of villous atrophy after a gluten challenge.

The three different Dutch patient cohorts described in this thesis did not contain any DQ2 and DQ8 negative patients. Based on the 6% occurrence of DQ2/8 negative patients reported by the European consortium, at least 10 such patients would be expected in our data set. Our inclusion criteria were not based on the presence of either DQ2 or DQ8, but only on the presence of a Marsh III lesion upon re-evaluation of the initial biopsy specimens. Patients presenting with less severe lesions (Marsh I or II) are frequently negative for both DQ2 and DQ8, indicating that DQ2 and DQ8 may be associated with more severe damage of the small intestine (Crusius et al., manuscript

submitted).[31] This would explain the non-occurrence of DQ2/8 negative patients in our cohorts, since only Marsh III patients were included and genotyped at the DQA1 and DQB1 loci. The DQ2/8 negative patients studied by the European consortium were not subjected to histological re-evaluation, and it is possible that some of the patients exhibited less severe Marsh I or II lesions.

Implications for population screening

Coeliac disease fulfils all five WHO criteria for population screening:[36] 1. It is a common disease, 2. Most patients are not recognized by clinical symptoms, 3. Highly sensitive and specific antibody screening tests are available, 4. An effective treatment is available, 5. Severe complications could occur in untreated patients. Certainly, several aspects of the disease plead in favour of population screening: the vast majority of patients remain unrecognised, untreated disease may lead to irreversible complications such as osteoporosis, autoimmune disorders and intestinal malignancies,[37-40] and increased mortality in untreated patients was observed.[40,41] One important aspect to consider is that the sensitivity of anti-Ema and anti-tTG antibodies is high in patients with (sub)total villous atrophy but much lower in patients with only partial villous atrophy.[42,43] A large proportion of patients may therefore be missed by antibody screening since patients presenting with partial villous atrophy are frequent. Another aspect to take into account is that although the gluten-free diet may be a very effective treatment for coeliac disease, it is not an easy diet. As asymptomatic patients do not experience clinical benefits, their motivation to adhere to the diet might be low. Indeed, it has been shown that dietary compliance of screening-detected patients was lower than compliance of symptomatic patients.[44] Furthermore, there is currently no evidence that treatment of asymptomatic patients would reduce morbidity and mortality.[45] These last two aspects would severely impact the clinical and cost benefits of population screening at this time. Pilot studies should be performed first to indicate whether population screening is likely to result in decreased morbidity and mortality in coeliac disease patients. Recently, a study investigating the costs and health benefits of population screening for coeliac disease in the Netherlands concluded that a screening programme was unlikely to be cost saving based on the direct medical costs of the complications, but that important health gains

could be achieved by prevention of these complications.[46] It was therefore recommended to start a large research project to further investigate the value of population screening.

If population screening would prove to be a cost-effective disease prevention strategy, who should be screened and how should this be done? Coeliac disease is rare in DQ2 and DQ8 negative individuals, so instead of screening the entire population a two-step strategy may be preferred. HLA typing for the presence of DQ2 and DQ8 genotypes will identify the approximately 30% of the general population that are potentially at risk for coeliac disease.[47,48] Only these individuals need to be subjected to serological screening for coeliac disease. But what should be done with individuals that carry only half of the DQ2 heterodimer? Since these individuals seem to have an increased risk of developing of the disease too, it may be justified to include them in the serological screening. However, this would increase the costs as more individuals have to be considered for serological testing. The strategy which may provide the best cost/benefit ratio is to screen for coeliac disease only in individuals that are homozygous DQA1*05-DQB1*02 (DR3/3) or heterozygous DQA1*05-DQB1*02/DQA1*0201-DQB1*02 (DR3/7). These genotypes were shown to be associated with the highest disease risk and together account for only 4% of the general population but for almost 60% of coeliac disease patients (see Chapter 8).

HLA genotyping could be included in the screening program for phenylketonuria (PKU) in which blood is collected from all newborns. One other advantage of population HLA typing for DQ2 and DQ8 is that these genotypes are also associated with an increased risk for other autoimmune disorders like type 1 diabetes mellitus, autoimmune thyroid disease and systemic lupus erythematosus.[16,49] Individuals that carry these risk genotypes could also be monitored for the development of other autoimmune disorders, thereby increasing the probability of early detection and prevention of disease associated complications. However, because of the conservative approach of population screening in the Netherlands, combined screening programmes for multiple disorders are unlikely to be approved.

The serological screening for coeliac disease should preferably be performed in early childhood, since the prevalence of other autoimmune disorders in coeliac disease patients was shown to be correlated with the duration of gluten intake.[37] Furthermore, compliance to the gluten-free diet is better in children diagnosed at a young age.[50]

However, patients may develop coeliac disease later in life, which means that the serological screenings may have to be repeated after a certain time.

Coeliac disease as a model for HLA-associated disorders

Coeliac disease is the only HLA-associated disorder for which both a major gene (HLA-DQ) and a major antigen (gluten peptides) are known. Association with the HLA region has been well established for many autoimmune diseases, but except for coeliac disease, the disease-causing gene is not known in any of these diseases. Genetic studies are limited in their resolving power to distinguish between genes, largely because of the extensive LD displayed by the HLA region. Therefore, functional studies are necessary to determine which of the genes are involved in the disease process. However, the affected tissue in most disorders is not easily accessible, which makes it almost impossible to perform these experiments. In contrast, small intestinal biopsies are routinely taken for diagnosis of coeliac disease and additional samples for research purposes can be obtained without significant additional risk or burden to the patient. This makes coeliac disease a unique model for studying the interaction between HLA molecules and their antigens, and the subsequent response by the immune system. These studies may lead to new insights into autoimmune destruction of tissues, which is of great importance for the understanding of autoimmune processes in general.

Contribution of non-HLA genes to coeliac disease

Non-HLA genes confer at least 50% of the genetic contribution

Although the HLA region seems essential for development of coeliac disease, it is not the sole genetic factor underlying this disease. The large difference in disease concordance rates between monozygous twins (86%) and HLA-identical siblings (30%) strongly suggests a contribution of non-HLA genes.[13,51] This is supported by the observation that <2% of HLA-DQ2 or DQ8 positive individuals eventually develop coeliac disease. Estimations of the relative contribution of the HLA region to the overall genetic risk vary from 21% to 44%, based on a multiplicative model.[14,52] The estimated contribution by the HLA region in the Dutch population is approximately 47%, based on the observed $\lambda_{s,\,HLA}$ of 4.6 and a λ_s of 25 for coeliac disease. (Chapter 2).[53] These results indicate that

the HLA contribution comprises at most half of the total genetic risk, and that non-HLA genes will contribute at least 50%.

The chance to find these non-HLA genes depends on the specific contribution of each of them. At the moment, ten whole genome screens have been performed in coeliac disease, including the two described in this thesis (Chapters 2 and 3).[53-61] While linkage to the HLA region was present in all of them, linkage to other regions reached much lower significance levels. These results indicate that the non-HLA contribution is very unlikely to be attributable to a single gene and, therefore, the contribution of each non-HLA locus is expected to be much smaller than that of the HLA region. Hence, the probability of detecting these non-HLA loci is lower and larger numbers of patients will be required to detect them. The total number of non-HLA susceptibility loci involved in coeliac disease will have to be elucidated by association and linkage studies involving large collections of independent cases and controls and families with multiple affected individuals.

Candidate gene studies

The search for susceptibility genes underlying complex diseases has in the past mainly been focussed on functional candidate genes, since information about localisation of the disease genes was not available. A logical approach is typing polymorphisms with a functional effect on the protein.[62] Association should be replicated in another data set from the same population, or in another population, before involvement of this gene can be accepted. However, this approach has not been very successful in complex diseases.[63,64] Failure to replicate association in a different data set may have several reasons, which are all due to typical features of complex diseases. Firstly, the second data set is a different sample, in which a different allele of the same gene may be associated with the disease (allelic heterogeneity, see also Chapter 1). Secondly, a locus associated with the disease in the first data set may only have a small effect in the second sample, especially when this sample is from a different population (locus heterogeneity). Thirdly, complex disease phenotypes can vary greatly between different patients. Defining the appropriate phenotype can be difficult, and the different data sets may contain different phenotypes, which may be associated with different genotypes.[64] Fourthly, associations may just occur by chance. To avoid these false-positive associations, higher significance

thresholds in the order of P $<10^{-6}$-10^{-8} have been proposed.[65] This would certainly dramatically reduce the false-positive rate, but it would also increase the false-negative rate. To reach such low P-values, cohorts of at least 1000 cases and controls are required, which can be difficult to collect.[65] As a compromise, association at a significance threshold of P <0.05, with independent confirmation in at least one different data set could be considered a more realistic proof of association of a given polymorphism with the disease.

Candidate gene studies in coeliac disease have mainly focussed on the gene encoding the cytotoxic T-lymphocyte-associated protein 4 (*CTLA4* at 2q33), because this gene has been implicated as a general susceptibility gene for autoimmunity (see also Chapter 6).[66] This gene contains only one non-silent missense SNP, the A/G polymorphism at position +49, and this polymorphism has therefore received most attention. Association was first reported in a case-control study in French patients and later confirmed by two studies in different Scandinavian populations.[67-69] The GG genotype conferred a slightly increased risk to coeliac disease in the Dutch population (Chapter 6). A microsatellite marker in the 3' untranslated region of the gene has also been tested, but association was not found.[70-72] Again, lack of association between a microsatellite marker and a disease might be attributed to the high mutation rate of the marker, resulting in the presence of several disease-associated alleles. Therefore, when testing a candidate gene, functional polymorphisms are the preferred type of polymorphism, although typing just one polymorphism is often not sufficient. LD can extend over long regions and the association pattern in the region surrounding a candidate gene must be determined before the causal variant can be identified. In a study that investigated the role of the *CTLA4* gene region in Graves' disease, type 1 diabetes mellitus and autoimmune hypothyroidism, LD was shown to extend over a region of 100kb, including the entire *CTLA4* gene and part of its neighbouring gene.[73] This LD block contained 78 SNPs, including the +49A/G polymorphism, of which no less than 55 were significantly associated with Graves' disease. Moreover, regression analysis indicated that the +49A/G polymorphism could not be the causal variant, which was mapped to a 6.1 kb region containing four other SNPs. The disease-susceptible haplotype of these SNPs was tested for mRNA expression and this haplotype was associated with lower levels of soluble *CTLA4* mRNA. Each SNP should be tested separately to

determine which of the four SNPs is actually responsible for the lowered expression. This study showed how a systematic approach of typing many SNPs in the entire candidate region resulted in mapping of the causal variant to a very small region and provide a functional explanation for the observed association with the risk haplotype.

Not many other candidate gene studies have been performed in coeliac disease. Two intragenic microsatellite markers in the genes encoding the functional candidates tissue transglutaminase (*TGM2* at 20q11.23) and interferon-γ (*IFNG* at 12q15) were tested in Dutch patients, but association could not be detected (Chapters 5 and 7, respectively).[74] Also, sequencing of the coding region of *TGM2* did not reveal differences between cases and controls.[75] Several studies investigated the contribution of the T cell receptor alpha (*TRA* at 14q11.2), beta (*TRB* at 7q34), gamma (*TRG* at 7p14.1) and delta (*TRD* at 14q11.2) genes, but no association was found with any of them.[76-78] Marginal associations with the mannose binding lectin gene (*MBL2* at 10q21.1) and the genes encoding the GM immunoglobulin allotypes (at 14q32.33) have been reported, but these results have to date not been confirmed in independent data sets.[79,80]

In conclusion, candidate gene analysis has not resulted in unambiguous identification of any non-HLA susceptibility gene for coeliac disease, except for a putative role of the *CTLA4* gene region. However, the contribution of the *CTLA4* region should be investigated by typing multiple SNPs in this region, in addition to the +49A/G polymorphism. Future studies should also include SNPs from the 6.1 kb region shown to harbour the causal variation for Graves' disease, type 1 diabetes mellitus and autoimmune hypothyroidism.

Localising non-HLA susceptibility genes

Candidate gene analysis can be a powerful tool for implicating genes in disease processes, but it has some serious limitations. A good understanding of the disease pathogenesis is essential to prioritise the list of possible genes. Furthermore, only genes with known functions can be considered. The completion of the Human Genome Project will eventually lead to identification and characterisation of all genes in the human genome. However, by July 2003, 40% of the 23,299 predicted protein coding genes in the Ensembl database were still of unknown function.

No genes conferring susceptibility to coeliac disease were known outside the HLA region at the beginning of this project. This meant that the only realistic approach was to map the position of susceptibility loci in the human genome by linkage analysis in families with multiple affected individuals. No large coeliac disease families were available and, therefore, a non-parametric linkage analysis approach using affected sibpairs was chosen. Affected sibpairs were preferred because these are more easily collected and genotyped compared to other relative pairs. However, it has been argued that mapping genes in more distant affected relative pairs might be more powerful.[81] A very important part of a linkage study is the phenotyping of the patients. Inclusion of patients with uncertain or different phenotypes can produce noise or even completely conceal a linkage signal.[82] Therefore, only sibpairs presenting with a proven Marsh III lesion were included in the study. Patients with less severe lesions were excluded, because it is currently unknown whether these lesions are caused by the same disease mechanisms.

To date, ten whole genome screens in search for coeliac disease loci have been completed: five were conducted in affected sibpairs,[53-57] three in extended families,[58,60,61] one in a population isolate[59] and one in a four-generation family (Chapter 3). Significant evidence for linkage was present in just two studies: in the population isolate[59] and in the Dutch affected sibpairs.[53] It is tempting to speculate that the significant linkage detected in our study was attributable to careful phenotyping of the patients. None of the other studies re-evaluated the biopsy specimens of the patients, although most of them thoroughly investigated the histological and clinical reports. However, this may be insufficient, since until just a few years ago coeliac disease was still quite rare in most countries, and pathologists may not have been experienced enough to make the correct diagnosis in damaged or poorly oriented biopsies. We were also lucky with the presence of a major non-HLA locus in the Dutch population. The data set used in the initial genome-wide screen had little power to detect loci with a $\lambda_s \leq 1.5$ and identification of such loci would have been difficult, even in a carefully phenotyped data set.

What have we gained from these ten linkage studies? As many as 50 non-HLA loci with nominal P-values ≤0.05 were found, thirteen of which were replicated in at least one other data set (Table 2). However, this has not yet lead to the identification of susceptibility genes in these regions. It is also apparent that none of the functional

candidate genes previously tested for association with coeliac disease are located in regions showing linkage to the disease (see previous section). A few candidate genes within the linked regions have been tested: the IL12B gene at 5qter, the natural killer cell immunoglobulin-like receptor (KIR) and leukocyte immunoglobulin-like receptor (LILR) gene clusters at 19q13.4, and the matrix metalloproteinase (MMP) 1 and 3 genes at 11qter.[83-85] However, no evidence for association between coeliac disease and any of them was found. Linkage to most regions only reached low significance, so they probably harbour genes with small effects, which would be difficult to identify by association analysis. Combining different data sets may result in data sets with sufficient power to identify these loci. The 5qter region seems the best candidate for this approach, as it is most consistent region between the populations. Also 4p15, 9p21 and 11qter are promising loci with replications in at least two studies.

Table 2. Overview of non-HLA loci reported in at least two linkage studies in coeliac disease with nominal P <0.05.

	Population[a]									
	Finland (sibpairs) (56)	Finland (isolate) (59)	Ireland (54)	Italy (55)	Netherlands (sibpairs) (Chapter 2)	Netherlands (family) (Chapter 3)	Northern Europe (58)	Sweden/ Norway (57)	UK (60)	USA (61)
Locus										
1p36	+								+	
3p24								+		+
4p15	+	+					+		+	
5qter	+	+	+	+	+			+		
9p21	+					+		+		
11p11			+						+	
11qter				+				+	+	
16q23						+			+	
17q21								+	+	
18q23									+	+
19p13.3			+				+			
19p13.1					+	+				
19q13.4			+						+	

[a] References are listed between brackets.

When no obvious candidate genes are present in a candidate region obtained by linkage analysis, systematic fine-mapping using linkage disequilibrium has to be performed to narrow-down the number of possible candidates. The remaining positional candidate genes all become equally interesting, independent of their putative functions. Only genes with expression patterns inconsistent with the disease may be excluded as possible disease susceptibility genes, for example those not expressed in small intestine in case of coeliac disease. None of the regions showing linkage to coeliac disease have been

subjected to fine-mapping yet, except the 19p13.1 locus in our study (Chapter 4 and next section).

Positional cloning of disease susceptibility genes

Linkage analysis is very suitable for roughly localising complex disease susceptibility genes in the genome. However, candidate regions obtained by linkage analysis tend to be quite large and generally contain many genes. This is due to the fact that affected relatives share large parts of their genomes, in case of sibpairs even 50%. Therefore, shared regions around the disease gene are also large, as they have not yet been subjected to recombination, resulting in large candidate regions obtained by linkage analysis. When no obvious candidate genes are present in the candidate region, fine-mapping has to be performed to narrow-down the region and thereby the number of possible candidate genes. For this purpose, an LD mapping strategy is applied, which involves the search for shared segments around the causal variant. It is based on the hypothesis that a mutation event occurred on a certain genetic background. Numerous recombinations have reduced the region around this variant, but small regions are still shared between carriers of the variant. Therefore, not only the causal variant will show association with the disease, but also surrounding polymorphisms. Consequently, covering the candidate region with densely spaced polymorphisms in unrelated patients, either in a case-control or TDT design, and testing for association of each polymorphism with the disease will pinpoint the location that contains the causal variant.

LD mapping will be most successful for detecting common disease-causing variants. Common variants have relatively high frequencies in the population and they are shared by a large proportion of patients, which allows for association analysis to detect them. Furthermore, these variants are also likely to be present in different populations, resulting in identical disease associated alleles in different populations.[86] On the other hand, multiple disease-causing mutations may have occurred in the same gene on different genetic backgrounds. This scenario results in the presence of multiple rare disease-causing variants, which are much harder to detect by association analysis.[87] Moreover, rare variants are more likely to be population-specific, with different disease-causing alleles in different populations. Isolated populations are most suitable for detecting rare variants as they are likely to display less allelic diversity.[88] Recently, LD

mapping following a linkage study proved to be successful for positional cloning of susceptibility genes in complex disorders. Common alleles of the *CAPN10* and *ADAM33* genes were shown to be associated with type 2 diabetes mellitus[89] and asthma[90], respectively. Different rare alleles of SNPs in the *CARD15/NOD2* gene were identified as susceptibility alleles for Crohn's disease.[91] However, the rare alleles of the associated SNPs were mostly present on a shared, common background, which facilitated their detection by LD mapping.[91,92] All three susceptibility genes were identified using a sequence-based map, in which mainly SNPs within genes were typed.

A slightly different strategy was used for fine-mapping of the candidate region for coeliac disease at 19p13.1. Our strategy was to use map-based polymorphisms (selected based on their relative positions) instead of sequence-based (selected based on their predicted effect on protein function).[93] This strategy was based on the observation that LD seems to have a block-like structure.[94] Therefore, one polymorphism in a haplotype block containing the causal variant would be sufficient to detect association. Recently, the haplotype block structure of chromosome 19 was published, showing that our candidate region resides in a region with low LD.[95] The advantage of low LD is that the shared region around the causal variant will be rather small and this will facilitate the identification of the actual disease associated gene. On the other hand, a higher density of polymorphisms may be necessary to detect the association. However, this information was not yet available when the fine-mapping study was initiated and, therefore, we started by typing all available microsatellite markers followed by equally spaced SNPs. The haplotype block information can be used for selecting additional SNPs in genes in the region 3' of *MYO9B*, by covering all haplotype blocks with at least one SNP and selecting more SNPs for regions with low LD. However, the haplotype block structure can differ substantially between populations so only information from closely related populations should be used.[82] The results obtained so far indicate that the causal variant is located within the *MYO9B* gene region. The alleles of the five SNPs in the *MYO9B* gene that were significantly increased in patients formed a haplotype that was also significantly increased in patients. Next, the disease associated SNPs have to be typed in the affected sibpair cohort to indicate whether the disease associated haplotype in the case-control cohort explains the linkage in this region, provided that the affected sibpair cohort has sufficient power for this purpose. Likewise, excess transmission of the disease associated

haplotype in the parent-case trios may confirm the association. Overall, the 19p13.1 region is the first to show systematic association in a coeliac disease candidate region identified by linkage analysis, and it provides the best chance so far for identification of a major non-HLA susceptibility gene for coeliac disease.

Our LD mapping strategy also involved a DNA pooling approach as a quick initial screening tool. This seemed applicable as 30% of the chromosomes were estimated to carry a disease associated allele (based on the assumption that 60% of patients carry a dominant risk allele at the 19p13.1 locus (see also Chapter 4)). Detection of association should be easily made in DNA pools, assuming that all patients carry the same disease-causing variant. However, association involving different rare variants may be missed. Indeed, the DNA pooling approach performed well, although some differences in significance levels were present between estimated and actual allele frequencies. One drawback of the DNA pooling approach is that only allele frequencies can be estimated. However, typing of single SNPs may not be sufficient to detect the association and haplotypes involving multiple SNPs will have to be constructed for which individual genotypes must be available. Therefore, DNA pooling can be applied as a quick tool for reducing the size of a candidate region, but single typing has to be performed within this smaller region to determine the disease-causing haplotype.

When different variants in the disease-causing haplotype are in complete LD, genetic studies cannot provide any more information about the identity of the actual disease variant. From that point onwards, RNA and protein studies have to be performed with the candidate genes within the region. For the coeliac disease locus on 19p13.1, mRNA expression of the remaining candidate genes could be investigated in small intestinal tissue of patients and normal controls. Differences in mRNA levels or length of the transcripts may lead to recognition of the gene involved and to elucidation of the disease-causing mechanism. Furthermore, antibodies against the candidate proteins may exist and they can be used for determining the localisation and expression level of the protein in the small intestine of patients and controls. Once differences at the mRNA or protein level have been detected, it is necessary to identify which variant is responsible for this and whether this variant is also part of the disease associated haplotype.

Future prospects

This project has provided important insights into the localisation of the major genes underlying coeliac disease in the Dutch population. Obviously, the next step is to identify and characterize the susceptibility gene in the 19p13.1 region. LD mapping has reduced the candidate region to approximately 150 kb containing only eight genes. Association analysis of additional SNPs in the genes 3' of *MYO9B* may reduce the region even further. The gene density in the 150 kb candidate region is very high, so the disease associated haplotype is likely to contain multiple genes. In that case, mRNA studies may provide more information about the identity of the disease-causing gene. When the disease-causing gene is identified, its function will have to be determined and also its role in the disease pathogenesis. This will provide important new information about the pathways involved in coeliac disease.

In addition, it is important to test whether the *MYO9B* region is also associated with coeliac disease in other populations. The Dutch population is not an isolated population so the disease-causing variant may also be carried by patients from other European or North American populations, although it may confer a lower risk. Combining the results of such studies would facilitate the estimation of the 19p13.1 locus contribution to coeliac disease in general.

The same LD mapping approach can be used for the candidate region at 6q21-22. However, the maximum lod score in this region was considerately lower than that of the 19p13.1 region. Therefore, it may be harder to detect association in the current case-control cohort. Increasing this cohort may be necessary to detect this locus. Another possibility is to include patients with type 1 diabetes mellitus as both diseases are likely to share part of their genetic background. The prevalence of coeliac disease is increased in patients with type 1 diabetes mellitus and linkage of this disease to 6q21 has been demonstrated in several data sets.[49,96] The presence of association between a polymorphism and both diseases would support involvement of that region in autoimmune processes, even if significance levels are low.

The genetic contribution to coeliac disease in Dutch patients cannot be explained completely by the action of the HLA-DQ, 19p13.1 and 6q21-22 loci. Several other genes with smaller genetic contributions are therefore expected to be involved as well. Promising loci are, for example, the *CTLA4* gene and the 5qter region. However, the

current case-control and parent-case cohorts are unlikely to have sufficient power to detect loci with such small effects. International collaboration with other groups or with the European coeliac disease consortium may therefore be required to obtain a complete understanding of the genetic contribution eventually.

And finally, the question remains whether the coeliac disease patients will derive any benefit from the results obtained by this project? A genome-wide screen is just the beginning of the search for non-HLA susceptibility genes, and identification of these genes was therefore not to be expected. However, the 19p13.1 candidate region provides a very good opportunity for identification of a major gene and the identity of this gene is likely to be known in the near future. This may not only point to a new pathway involved in coeliac disease, but it may also provide new targets for possible therapeutic intervention. Furthermore, genotyping of the 19p13.1 disease-causing variant may assist in diagnosis of the disease, for example in patients presenting with minor lesions or those who are already on a gluten-free diet. In addition, it is important to know whether this gene confers different risks in DR3/3, 3/7 and 3/X individuals. This information could then be applied to design family and population screening strategies for the identification of individuals at high risk of developing coeliac disease. Such individuals could be subjected to regular clinical investigations in order to diagnose the disease at an early stage.

References

1. Csizmadia CG, Mearin ML, von Blomberg BM, Brand R, Verloove-Vanhorick SP. An iceberg of childhood coeliac disease in the Netherlands. *Lancet* 1999; **353:** 813-4.
2. Rostami K, Mulder CJ, Werre JM*, et al.* High prevalence of celiac disease in apparently healthy blood donors suggests a high prevalence of undiagnosed celiac disease in the Dutch population. *Scand J Gastroenterol* 1999; **34:** 276-9.
3. Nederlandse Coeliakie Vereniging (www.coeliakievereniging.nl).
4. Maki M, Holm K, Lipsanen V*, et al.* Serological markers and HLA genes among healthy first-degree relatives of patients with coeliac disease. *Lancet* 1991; **338:** 1350-3.
5. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 1992; **102:** 330-54.
6. Stenhammar L, Brandt A, Wagermark J. A family study of coeliac disease. *Acta Paediatr Scand* 1982; **71:** 625-8.
7. Rolles CJ, Myint TO, Sin WK, Anderson M. Proceedings: Family study of coeliac disease. *Gut* 1974; **15:** 827.
8. Stevens FM, Lloyd R, Egan-Mitchell B*, et al.* Reticulin antibodies in patients with coeliac disease and their relatives. *Gut* 1975; **16:** 598-602.

9. Stokes PL, Ferguson R, Holmes GK, Cooke WT. Familial aspects of coeliac disease. *Q J Med* 1976; **45:** 567-82.
10. Vogelsang H, Wyatt J, Penner E, Lochs H. Screening for celiac disease in first-degree relatives of patients with celiac disease by lactulose/mannitol test. *Am J Gastroenterol* 1995; **90:** 1838-42.
11. Mylotte M, Egan-Mitchell B, Fottrell PF, McNicholl B, McCarthy CF. Family studies in coeliac disease. *Q J Med* 1974; **43:** 359-69.
12. David TJ, Ajdukiewicz AB. A family study of coeliac disease. *J Med Genet* 1975; **12:** 79-82.
13. Sollid LM, Thorsby E. HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 1993; **105:** 910-22.
14. Petronzelli F, Bonamico M, Ferrante P*, et al.* Genetic contribution of the HLA region to the familial clustering of coeliac disease. *Ann Hum Genet* 1997; **61:** 307-17.
15. Alper CA, Fleischnick E, Awdeh Z, Katz AJ, Yunis EJ. Extended major histocompatibility complex haplotypes in patients with gluten-sensitive enteropathy. *J Clin Invest* 1987; **79:** 251-6.
16. Candore G, Lio D, Colonna Romano G, Caruso C. Pathogenesis of autoimmune diseases associated with 8.1 ancestral haplotype: effect of multiple gene interactions. *Autoimmun Rev* 2002; **1:** 29-35.
17. Bolognesi E, Dalfonso S, Rolando V, Fasano ME, Pratico L, Momigliano-Richiardi P. MICA and MICB microsatellite alleles in HLA extended haplotypes. *Eur J Immunogenet* 2001; **28:** 523-30.
18. Louka AS, Moodie SJ, Karell K*, et al.* A collaborative European search for non-DQA1*05-DQB1*02 celiac disease loci on HLA-DR3 haplotypes: analysis of transmission from homozygous parents. *Hum Immunol* 2003; **64:** 350-8.
19. Louka AS, Lie BA, Talseth B*, et al.* Coeliac disease patients carry conserved HLA-DR3-DQ2 haplotypes revealed by association of TNF alleles. *Immunogenetics* 2003; **55:** 339-43.
20. Bolognesi E, Karell K, Percopo S*, et al.* Additional factor in some HLA DR3/DQ2 haplotypes confers a fourfold increased genetic risk of celiac disease. *Tissue Antigens* 2003; **61:** 308-16.
21. McManus R, Wilson AG, Mansfield J, Weir DG, Duff GW, Kelleher D. TNF2, a polymorphism of the tumour necrosis-alpha gene promoter, is a component of the celiac disease major histocompatibility complex haplotype. *Eur J Immunol* 1996; **26:** 2113-8.
22. de la Concha EG, Fernandez-Arquero M, Vigil P*, et al.* Celiac disease and TNF promoter polymorphisms. *Hum Immunol* 2000; **61:** 513-7.
23. Garrote JA, Arranz E, Telleria JJ, Castro J, Calvo C, Blanco-Quiros A. TNF alpha and LT alpha gene polymorphisms as additional markers of celiac disease susceptibility in a DQ2-positive population. *Immunogenetics* 2002; **54:** 551-5.
24. Congia M, Frau F, Lampis R*, et al.* A high frequency of the A30, B18, DR3, DRw52, DQw2 extended haplotype in Sardinian celiac disease patients: further evidence that disease susceptibility is conferred by DQ A1*0501, B1*0201. *Tissue Antigens* 1992; **39:** 78-83.
25. Grimaldi MC, Clayton J, Pontarotti P, Cambon-Thomsen A, Crouau-Roy B. New highly polymorphic microsatellite marker in linkage disequilibrium with HLA-B. *Hum Immunol* 1996; **51:** 89-94.
26. Kaur G, Sarkar N, Bhatnagar S*, et al.* Pediatric celiac disease in India is associated with multiple DR3-DQ2 haplotypes. *Hum Immunol* 2002; **63:** 677-82.
27. Mearin ML, Bouquet J, Mourad N*, et al.* HLA-DR antigens and phenotypes in Dutch coeliac children and their families. *Clin Genet* 1985; **27:** 45-50.
28. Ploski R, Ek J, Thorsby E, Sollid LM. 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* 1993; **41:** 173-7.

29. Louka AS, Nilsson S, Olsson M*, et al.* HLA in coeliac disease families: a novel test of risk modification by the 'other' haplotype when at least one DQA1*05-DQB1*02 haplotype is carried. *Tissue Antigens* 2002; **60:** 147-54.
30. Karell K, Louka AS, Moodie SJ*, et al.* Hla types in celiac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the european genetics cluster on celiac disease. *Hum Immunol* 2003; **64:** 469-77.
31. Wahab PJ, Crusius JB, Meijer JW, Mulder CJ. Gluten challenge in borderline gluten-sensitive enteropathy. *Am J Gastroenterol* 2001; **96:** 1464-9.
32. Tursi A, Brandimarte G. The symptomatic and histologic response to a gluten-free diet in patients with borderline enteropathy. *J Clin Gastroenterol* 2003; **36:** 13-7.
33. Farrell RJ, Kelly CP. Diagnosis of celiac sprue. *Am J Gastroenterol* 2001; **96:** 3237-46.
34. Mulder CJ. When is a coeliac a coeliac? Report of a working group of the United European Gastroenterology Week in Amsterdam, 2001. *Eur J Gastroenterol Hepatol* 2001; **13:** 1123-8.
35. Walker-Smith JA. Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. *Arch Dis Child* 1990; **65:** 909-11.
36. Fasano A. European and North American populations should be screened for coeliac disease. *GUT* 2003; **52:** 168-9.
37. Ventura A, Magazzu G, Greco L. Duration of exposure to gluten and risk for autoimmune disorders in patients with celiac disease. *Gastroenterology* 1999; **117:** 297-303.
38. Cellier C, Flobert C, Cormier C, Roux C, Schmitz J. Severe osteopenia in symptom-free adults with a childhood diagnosis of coeliac disease. *Lancet* 2000; **355:** 806.
39. Holmes GK, Prior P, Lane MR, Pope D, Allan RN. Malignancy in coeliac disease--effect of a gluten free diet. *Gut* 1989; **30:** 333-8.
40. Corrao G, Corazza GR, Bagnardi V*, et al.* Mortality in patients with coeliac disease and their relatives: a cohort study. *Lancet* 2001; **358:** 356-61.
41. Logan RF, Rifkind EA, Turner ID, Ferguson A. Mortality in celiac disease. *Gastroenterology* 1989; **97:** 265-71.
42. Rostami K, Kerckhaert J, Tiemessen R, von Blomberg BM, Meijer JW, Mulder CJ. Sensitivity of antiendomysium and antigliadin antibodies in untreated celiac disease: disappointing in clinical practice. *Am J Gastroenterol* 1999; **94:** 888-94.
43. Tursi A, Brandimarte G, Giorgetti GM. Prevalence of antitissue transglutaminase antibodies in different degrees of intestinal damage in celiac disease. *Clin Gastroenterol* 2003; **36:** 219-21.
44. Fabiani E, Taccari LM, Ratsch IM, Di Giuseppe S, Coppa GV, Catassi C. Compliance with gluten-free diet in adolescents with screening-detected celiac disease: a 5-year follow-up study. *J Pediatr* 2000; **136:** 841-3.
45. Kumar P. European and North American populations should be screened for coeliac disease. *GUT* 2003; **52:** 170-1.
46. Sneeuw KCA, Brodin H, Broker M*, et al.* Screening for celiac disease in the Netherlands: estimation of costs and benefits. *TNO report* 2001.
47. Moghaddam PH, de Knijff P, Schipper RF*, et al.* Selective co-evolution of the D6STNFa microsatellite region with HLA class I and II loci. *Tissue Antigens* 1998; **52:** 213-9.
48. Schipper RF, Schreuder GM, D'Amaro J, Oudshoorn M. HLA gene and haplotype frequencies in Dutch blood donors. *Tissue Antigens* 1996; **48:** 562-74.
49. Pociot F, McDermott MF. Genetics of type 1 diabetes mellitus. *Genes Immun* 2002; **3:** 235-49.
50. Hogberg L, Grodzinsky E, Stenhammar L. Better dietary compliance in patients with coeliac disease diagnosed in early childhood. *Scand J Gastroenterol* 2003; **38:** 751-4.

51. Greco L, Romino R, Coto I*, et al.* The first large population based twin study of coeliac disease. *Gut* 2002; **50:** 624-8.
52. Bevan S, Popat S, Braegger CP*, et al.* Contribution of the MHC region to the familial risk of coeliac disease. *J Med Genet* 1999; **36:** 687-90.
53. van Belzen MJ, Meijer JW, Sandkuijl LA*, et al.* A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* 2003; **125:** 1032-41.
54. Zhong F, McCombs CC, Olson JM*, et al.* An autosomal screen for genes that predispose to celiac disease in the western counties of Ireland. *Nat Genet* 1996; **14:** 329-33.
55. Greco L, Corazza G, Babron MC*, et al.* Genome search in celiac disease. *Am J Hum Genet* 1998; **62:** 669-75.
56. Liu J, Juo SH, Holopainen P*, et al.* Genomewide linkage analysis of celiac disease in Finnish families. *Am J Hum Genet* 2002; **70:** 51-9.
57. Naluai AT, Nilsson S, Gudjonsdottir AH*, et al.* 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* 2001; **9:** 938-44.
58. Popat S, Bevan S, Braegger CP*, et al.* Genome screening of coeliac disease. *J Med Genet* 2002; **39:** 328-31.
59. Woolley N, Holopainen P, Ollikainen V*, et al.* A new locus for coeliac disease mapped to chromosome 15 in a population isolate. *Hum Genet* 2002; **111:** 40-5.
60. King AL, Yiannakou JY, Brett PM*, et al.* A genome-wide family-based linkage study of coeliac disease. *Ann Hum Genet* 2000; **64:** 479-90.
61. Neuhausen SL, Feolo M, Camp NJ, Farnham J, Book L, Zone JJ. Genome-wide linkage analysis for celiac disease in North American families. *Am J Med Genet* 2002; **111:** 1-9.
62. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994; **265:** 2037-48.
63. Glazier AM, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. *Science* 2002; **298:** 2345-9.
64. Tabor HK, Risch NJ, Myers RM. Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 2002; **3:** 391-7.
65. Todd JA. Interpretation of results from genetic studies of multifactorial diseases. *Lancet* 1999; **354 Suppl 1:** SI15-6.
66. Kristiansen OP, Larsen ZM, Pociot F. CTLA-4 in autoimmune diseases--a general susceptibility gene to autoimmunity? *Genes Immun* 2000; **1:** 170-84.
67. Djilali-Saiah I, Schmitz J, Harfouch-Hammoud E, Mougenot JF, Bach JF, Caillat-Zucman S. CTLA-4 gene polymorphism is associated with predisposition to coeliac disease. *Gut* 1998; **43:** 187-9.
68. Naluai AT, Nilsson S, Samuelsson L*, et al.* 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* 2000; **56:** 350-5.
69. Popat S, Hearle N, Wixey J*, et al.* Analysis of the CTLA4 gene in Swedish coeliac disease patients. *Scand J Gastroenterol* 2002; **37:** 28-31.
70. King AL, Moodie SJ, Fraser JS*, et al.* CTLA-4/CD28 gene region is associated with genetic susceptibility to coeliac disease in UK families. *J Med Genet* 2002; **39:** 51-4.
71. Holopainen P, Arvas M, Sistonen P*, et al.* CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to celiac disease. A linkage and family-based association study. *Tissue Antigens* 1999; **53:** 470-5.
72. Popat S, Hearle N, Hogberg L*, et al.* Variation in the CTLA4/CD28 gene region confers an increased risk of coeliac disease. *Ann Hum Genet* 2002; **66:** 125-37.
73. Ueda H, Howson JM, Esposito L*, et al.* Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 2003; **30:** 30.

74. van Belzen MJ, Mulder CJ, Pearson PL, Houwen RH, Wijmenga C. The tissue transglutaminase gene is not a primary factor predisposing to celiac disease. *Am J Gastroenterol* 2001; **96:** 3337-40.
75. Aldersley MA, Hamlin PJ, Jones PF, Markham AF, Robinson PA, Howdle PD. No polymorphism in the tissue transglutaminase gene detected in coeliac disease patients. *Scand J Gastroenterol* 2000; **35:** 61-3.
76. Roschmann E, Wienker TF, Gerok W, Volk BA. T-cell receptor variable genes and genetic susceptibility to celiac disease: an association and linkage study. *Gastroenterology* 1993; **105:** 1790-6.
77. Roschmann E, Wienker TF, Volk BA. Role of T cell receptor delta gene in susceptibility to celiac disease. *J Mol Med* 1996; **74:** 93-8.
78. Yiannakou JY, Brett PM, Morris MA*, et al.* Family linkage study of the T-cell receptor genes in coeliac disease. *Ital J Gastroenterol Hepatol* 1999; **31:** 198-201.
79. Bouguerra F, Dugoujon JM, Babron MC*, et al.* Susceptibility to coeliac disease in Tunisian children and GM immunoglobulin allotypes. *Eur J Immunogenet* 1999; **26:** 293-7.
80. Boniotto M, Braida L, Spano A*, et al.* Variant mannose-binding lectin alleles are associated with celiac disease. *Immunogenetics* 2002; **54:** 596-8.
81. Risch N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 1990; **46:** 229-41.
82. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 2003; **33 Suppl:** 228-37.
83. Louka AS, Torinsson Naluai A, D'Alfonso S*, et al.* The IL12B gene does not confer susceptibility to coeliac disease. *Tissue Antigens* 2002; **59:** 70-2.
84. Louka AS, Stensby EK, Ek J, Gudjonsdottir AH, Ascher H, Sollid LM. Coeliac disease candidate genes: no association with functional polymorphisms in matrix metalloproteinase 1 and 3 gene promoters. *Scand J Gastroenterol* 2002; **37:** 931-5.
85. Moodie SJ, Norman PJ, King AL*, et al.* Analysis of candidate genes on chromosome 19 in coeliac disease: an association study of the KIR and LILR gene clusters. *Eur J Immunogenet* 2002; **29:** 287-91.
86. Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000; **405:** 847-56.
87. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001; **69:** 124-37.
88. Wright AF, Carothers AD, Pirastu M. Population choice in mapping genes for complex diseases. *Nat Genet* 1999; **23:** 397-404.
89. Horikawa Y, Oda N, Cox NJ*, et al.* Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 2000; **26:** 163-75.
90. Van Eerdewegh P, Little RD, Dupuis J*, et al.* Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* 2002; **418:** 426-30.
91. Hugot JP, Chamaillard M, Zouali H*, et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **411:** 599-603.
92. Vermeire S, Wild G, Kocher K*, et al.* CARD15 genetic variation in a Quebec population: prevalence, genotype-phenotype relationship, and haplotype structure. *Am J Hum Genet* 2002; **71:** 74-83.
93. Peltonen L, McKusick VA. Genomics and medicine. Dissecting human disease in the postgenomic era. *Science* 2001; **291:** 1224-9.
94. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001; **29:** 229-32.
95. Phillips MS, Lawrence R, Sachidanandam R*, et al.* Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nat Genet* 2003; **33:** 382-7.

96. Holmes GK. Screening for coeliac disease in type 1 diabetes. *Arch Dis Child* 2002; **87:** 495-8.

Summary

Coeliac disease is a common food intolerance with a complex genetic aetiology. It is caused by ingestion of gluten peptides from wheat and related proteins from barley and rye in genetically susceptible individuals. The disease affects the small intestine and leads to abnormalities ranging from the infiltration of the villous epithelium by lymphocytes to total villous atrophy. Clinical symptoms include gastrointestinal complaints such as diarrhoea and abdominal pain but also fatigue, weight loss, anaemia, osteopenia, growth retardation and failure to thrive. Coeliac disease is treated by a life-long gluten-free diet; there is no drug therapy available. Complete understanding of the genetic factors underlying the disease will provide insight into the pathways involved in the disease and this will hopefully identify possible targets for development of a therapy.

One important genetic factor contributing to coeliac disease is the human leukocyte antigen (HLA) DQ locus. The majority of patients are positive for DQ2, and almost all of the remaining patients are positive for DQ8. However, the contribution of the HLA region comprises at most half of the total genetic risk. Therefore, non-HLA genes must also play an important role in coeliac disease pathogenesis, but little is known about the location and identity of these genes. The results described in this thesis provide important new insights into the genetic background of coeliac disease in general and into the specific genetic factors underlying the disease in the Dutch population.

Chapter 2 describes the first localisation of susceptibility loci for coeliac disease in the Dutch population. A genome-wide screen was performed in a well-characterized set of affected sibpairs. Linkage analysis revealed the presence of two important non-HLA loci conferring a considerable risk to coeliac disease. The major locus was found to be located at chromosome region 19p13.1, which was the first to reach genome-wide significance in coeliac disease in an outbred population. The other susceptibility locus is located at chromosome region 6q21-22 and reached the threshold for genome-wide suggestive linkage. This region is also implicated in other autoimmune disorders and it may therefore harbour a general susceptibility gene for autoimmunity. Both the 19p13.1 and 6q21-22 loci present novel susceptibility loci for coeliac disease. A second genome-wide screen was performed in a four-generation family with 17 coeliac disease patients. The results, described in **Chapter 3**, showed that neither the chromosome 19p13.1 locus nor the chromosome 6q21-22 locus contributed significantly to the disease. Surprisingly, a third locus, located at chromosome region 9p21-13, was identified as the major locus in

this family. This locus had also been implicated in Scandinavian families with coeliac disease, and probably presents a locus with a small risk to coeliac disease in general. These results show that a single large family can provide a unique opportunity for mapping complex disease susceptibility genes with small effects in the general patient population.

Chapter 4 describes the systematic fine-mapping of the 19p13.1 candidate region, which has not been performed in any of the candidate regions found in other linkage studies in coeliac disease. The initial candidate region of 3 Mb, which contained 92 genes, was narrowed-down to 450 kb and only 12 genes by association analysis using microsatellite markers. Subsequent typing of single nucleotide polymorphisms (SNPs) in this region pinpointed the location of the gene to the final 150 kb of this region, with eight possible candidate genes left. Further research is necessary to determine the extent of the associated region and the identity of the disease-causing gene. The 19p13.1 region is the first candidate region identified by linkage analysis in which association is also present, providing the best opportunity so far for identifying a non-HLA susceptibility gene for coeliac disease.

Three functional candidate genes, located outside the two regions showing linkage to coeliac disease in the affected sibpairs, were tested for association (**Chapters 5-7**). The enzyme tissue transglutaminase modifies gluten peptides into epitopes with strong affinity for HLA-DQ2 and DQ8, resulting in a greatly enhanced T cell response. The gene encoding tissue transglutaminase (*TGM2*) was tested for association with coeliac disease and was excluded as primary factor in coeliac disease pathogenesis, as described in **Chapter 5**. The cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) has been implicated as a general susceptibility gene for autoimmunity. **Chapter 6** describes the association analysis of the functional *CTLA4* +49A/G polymorphism in a cohort of Dutch patients and controls. The *CTLA4* gene was also excluded as a major susceptibility gene for coeliac disease, although a small effect of the homozygous GG genotype was observed. The cytokine interferon-γ (IFN-γ) plays an important role in the pathogenesis of coeliac disease. The CA-repeat in the *IFNG* gene, of which a certain allele was shown to be associated with high expression of IFN-γ, was tested for

association. Again, no evidence for a major contribution of this gene to coeliac disease risk was found, as described in **Chapter 7**.

Finally, the contribution of the HLA region was evaluated and this is described in **Chapter 8**. An extensive scan of the entire HLA region was performed in simplex coeliac disease and control families. Phase-known DQ2-positive haplotypes from patients were compared to control DQ2 haplotypes to determine whether there was evidence for the presence of an additional HLA risk locus. These results indicated that the contribution of the HLA region is mainly attributable to HLA-DQ2 and that a significant contribution of other HLA genes is unlikely. In addition, individuals homozygous for DQ2 or heterozygous for DQA1*05-DQB1*02/DQA1*0201-DQB1*02 were found to be at five-fold increased risk for development of coeliac disease. This risk is conferred by the presence of a second DQB1*02 allele next to a DQA1*05-DQB1*02 haplotype, and seemed to be independent of the second DQA1 allele.

Samenvatting

Coeliakie (spreek uit: *seuliakíe*) is een veelvoorkomende voedselintolerantie met een complexe genetische achtergrond. Het wordt veroorzaakt door het eten van gluten eiwitten die voorkomen in tarwe, gerst en rogge. De ziekte tast het slijmvlies van de dunne darm aan, met afwijkingen variërend van de infiltratie van het darmepitheel door lymfocyten tot en met het totaal verdwijnen van de darmvlokken. Deze schade aan de dunne darm kan leiden tot diarree en buikpijn, maar ook tot chronische vermoeidheid, vermagering, bloedarmoede, botontkalking en een kort postuur. Er bestaat geen medicijn voor de behandeling van coeliakie en daarom moeten coeliakiepatiënten de rest van hun leven een strikt glutenvrij dieet volgen. Begrip van de genetische factoren die coeliakie veroorzaken zal meer inzicht geven in de processen die betrokken zijn bij de ziekte en dit zal hopelijk leiden tot nieuwe behandelingsmogelijkheden.

Een belangrijke genetische factor die bijdraagt aan coeliakie is het "human leukocyte antigen" (HLA) DQ locus. Een grote meerderheid van de coeliakiepatiënten draagt hiervan een bepaalde vorm: DQ2. Vrijwel alle DQ2-negatieve patiënten zijn positief voor DQ8. De bijdrage van het HLA gebied aan de totale genetische achtergrond van coeliakie bedraagt echter maximaal 50%. Daarom moeten er ook niet-HLA genen bij coeliakie betrokken zijn, maar er was vrijwel niets over deze genen bekend bij aanvang van dit onderzoek. De in dit proefschrift beschreven resultaten leveren een belangrijke bijdrage aan het inzicht in de genetische factoren die coeliakie veroorzaken, niet alleen voor de Nederlandse patiënten maar ook voor coeliakie in het algemeen.

In **hoofdstuk 2** wordt de localisatie van de genen die betrokken zijn bij coeliakie in de Nederlandse bevolking beschreven. Een dergelijk onderzoek was nog nooit in Nederland uitgevoerd. Hiervoor werd een hele genoom screen uitgevoerd bij broers en zussen met bewezen coeliakie ("aangedane sibparen"). Broers en zussen delen gemiddeld 50% van hun erfelijk materiaal, maar als ze beiden coeliakie hebben is het te verwachten dat ze 100% van de coeliakie veroorzakende genen zullen delen. Het gebied rondom zo'n gen zal ook gedeeld worden tussen de aangedane sibparen en dit gegeven wordt gebruikt om de coeliakie veroorzakende genen te localiseren. Een hele genoom screen houdt in dat het gehele erfelijke materiaal (genoom) op ongeveer 300 posities, die op een vaste afstand van elkaar liggen, wordt onderzocht. Op elke positie wordt dan bekeken of de aangedane sibparen meer dan de verwachte 50% van het DNA delen, wat een aanwijzing is dat daar een gen dat coeliakie veroorzaakt moet liggen. In de Nederlandse bevolking

blijken naast het HLA-DQ locus nog twee belangrijke genen voor te komen die beiden een aanzienlijke bijdrage leveren aan het risico op het ontwikkelen van coeliakie. Het eerste gen ligt op chromosoom 19 en het bewijs voor de aanwezigheid van een coeliakie veroorzakend gen in dit gebied is veel sterker dan voor enig ander gebied dat werd gevonden door één van de zeven andere genoom screens die wereldwijd zijn uitgevoerd bij families met coeliakie. Het tweede gen ligt op chromosoom 6 (buiten het HLA gebied) in een gebied dat ook betrokken is bij andere auto-immuun ziekten en het is daarom mogelijk dat dit gen betrokken is bij auto-immuniteit in het algemeen. Een tweede hele genoom screen werd uitgevoerd in een familie met 17 coeliakiepatiënten in vier generaties. Uit de resultaten, die in **hoofdstuk 3** zijn beschreven, blijkt dat de genen op chromosoom 19 en 6 geen grote bijdrage leveren aan de ontwikkeling van coeliakie in deze uitzonderlijke familie. Het belangrijkste gen bleek op chromosoom 9 te liggen en dit gebied is ook betrokken bij coeliakie in Scandinavische coeliakiepatiënten. Dit gebied bevat waarschijnlijk een gen met een kleine bijdrage aan de ontwikkeling van coeliakie in de algemene bevolking.

In **hoofdstuk 4** wordt de vervolgstudie van het chromosoom 19 gebied beschreven. Na de genoom screen was het kandidaatgebied nog behoorlijk groot en bevatte 92 genen, wat teveel is om allemaal te testen. Door middel van associatie analyse kon de kandidaatregio worden teruggebracht tot een gebied dat slechts acht genen bevatte. Verder onderzoek is noodzakelijk om aan te kunnen tonen welk van deze acht genen nu uiteindelijk coeliakie veroorzaakt. Het inperken van een kandidaatgebied voor coeliakie door middel van associatie analyse was tot nu toe nog nooit uitgevoerd. De behaalde resultaten bieden dan ook de beste kans ooit op het te identificeren van een belangrijk niet-HLA gen dat coeliakie veroorzaakt.

Drie functionele kandidaatgenen, die niet op chromosoom 6, 9 of 19 liggen, werden getest op hun mogelijke betrokkenheid bij coeliakie (**hoofdstukken 5 t/m 7**). Het enzym tissue transglutaminase kan gluten omzetten in een vorm die goed gebonden wordt door HLA-DQ2 en DQ8 eitwitten, wat leidt tot een sterk verhoogde T-cel reactie. Het gen dat codeert voor tissue transglutaminase werd onderzocht op associatie met coeliakie, wat beschreven wordt in **hoofdstuk 5**. Echter, het bleek dat dit gen kon worden uitgesloten als een gen dat coeliakie veroozaakt. Het gen dat codeert voor het "cytotoxic T-lymphocyte-associated" eiwit 4 (CTLA4) is betrokken bij verschillende

andere auto-immuun ziekten en is daarom ook een aantrekkelijke kandidaat voor coeliakie. Helaas bleek ook het *CTLA4* gen geen grote rol te spelen in de ontwikkeling van coeliakie, wat beschreven wordt in **hoofdstuk 6**. Interferon-γ speelt in belangrijke rol bij de ontwikkeling van de schade aan de dunne darm bij coeliakie. Daarom werd ook het gen dat codeert voor interferon-γ getest op associatie met coeliakie en in het bijzonder één bepaalde variant die met hoge expressie van het interferon-γ eiwit geassocieerd is. In **hoofdstuk 7** worden de resultaten beschreven en het bleek dat ook het interferon-γ gen niet betrokken is bij het ontstaan van coeliakie.

Tot slot werd de bijdrage van het HLA gebied onderzocht, wat beschreven is in **hoofdstuk 8**. Er werd een uitgebreid onderzoek van het HLA gebied uitgevoerd in families met één kind met coeliakie en in controle families met één kind zonder coeliakie. Het doel was te onderzoeken of er nóg een HLA gen betrokken is bij coeliakie, naast HLA-DQ2, maar hier kon geen bewijs voor worden gevonden. Wel bleek dat mensen die twee kopiën van DQ2 bezitten, of één DQ2 kopie plus een extra DQB1*02 variant, een vijf keer grotere kans hebben om coeliakie te krijgen dan mensen die maar één kopie van DQ2 bezitten.

Dankwoord

Het zit erop! Rest mij alleen het dankwoord nog. Want als zo'n boekje af is realiseer je je pas goed hoeveel mensen hebben bijgedragen aan de totstandkoming ervan.

Allereerst wil ik alle patiënten en hun familieleden die hebben deelgenomen aan dit onderzoek bedanken. Zonder jullie had dit boekje er nooit gelegen. De Nederlandse Coeliakie Vereniging (NCV) is van essentieël belang gebleken voor het bereiken van families met "aangedane sibparen" voor de linkage studie (hoofdstuk 2). Ik wil de NCV dan ook graag bedanken voor de mogelijkheid om tijdens patiëntendagen mijn project te presenteren en via het kwartaalblad de resultaten van het onderzoek bekend te maken.

Cisca, jij bent voor mij toch wel de belangrijkste persoon geweest die bij dit project betrokken was. Jij hebt het aangedurfd mij aan te nemen terwijl ik nog nooit van het woord "coeliakie" gehoord had, niet wist wat al die medische termen betekenden en ook nog eens geen kaas had gegeten van genetica. Heel erg bedankt voor alles. Ik vind het heel leuk dat je nu ook nog mijn promotor kunt zijn!

Peter, bedankt voor het vertrouwen dat jij in mij hebt gesteld en de vrijheid die je mij gegeven hebt. Roderick, bedankt voor je kritische blik op de klinische aspecten van het project en de papers.

Winny, zonder jou was het waarschijnlijk niet gelukt al die patiënten te verzamelen en de biopten op te vragen. Vele uren heb jij aan de telefoon gezeten om potentiële deelnemers enthousiast te maken voor het onderzoek. Ik heb echt enorm kunnen lachen om de verhalen die je met me deelde. Heel erg bedankt voor je inzet en je hulp. En natuurlijk ben ik ook Anath en Arjenne zeer dankbaar voor het verzamelen van de eerste families, zelfs al voordat ik aan dit project begon.

Alfons, ik kan je niet genoeg bedanken voor al het werk dat je mij uit handen hebt genomen. Waarschijnlijk zullen we later uitgelachen worden als we zeggen dat we een hele genoom screen met de hand hebben uitgevoerd, maar ik ben er trots op dat het ons is gelukt. Ik ben erg blij dat je me ook nog wilt bijstaan tijdens de promotie als paranimf.

Jackie, jij bent echt goud waard! Dank je wel voor het corrigeren van het Engels in dit hele boekje en alle papers, en voor het regelen van zovele administratieve taken. Dankzij jou was ik nog net op tijd met het indienen van de papieren voor mijn promotie,

bedankt! Harry, bedankt voor het invoeren van alle patiëntgegevens in de database, waardoor er structuur in die enorme hoeveelheid data kwam.

Lodewijk Sandkuijl heeft mij geleerd hoe een degelijk onderzoek in de humane genetica uitgevoerd moet worden. Dankzij hem ben ik me bewust geworden van het belang van nauwkeurig fenotyperen, genotyperen en analyseren, en ik zal deze kennis ook in de toekomst blijven toepassen. Ik vind het heel jammer dat hij, na ruim vier jaar betrokken te zijn geweest bij dit onderzoek, de successen niet meer heeft mogen meemaken.

Bart en Jonathan, ik vond het leuk tussen jullie in te zitten tijdens het schrijven van jullie proefschriften. Toen werd mij ook duidelijk dat ik mijn boekje gewoon in Word zou gaan opmaken! Dineke, Begoña en Dalila, dank voor jullie komst op de afdeling, om wat tegenwicht te geven tegen dat "mannenwereldje". Steven, ik heb genoten van de vele discussies over de opzet en analyse van onze studies en nog bedankt voor het gebruik van je "spreadsheetjes"! Eric en Ruben, jullie ben ik zeer dankbaar voor het isoleren van de bijna 2000 DNA-tjes voor mijn project! Sasha, bedankt voor alle SNPs die je getypeerd hebt. Mijn studenten Helena, Arantza en Martine wil ik bedanken voor hun inzet. Jullie bijdragen zijn terug te vinden in de hoofdstukken 8, 7 en 3! Veel succes met jullie eigen promotie-onderzoek. En Bobby, jij bedankt voor je bijdrage aan de statistiek van de associatie analyses en het controleren van mijn analyses.

Chris Mulder, bedankt voor de onuitputtelijke stroom patiënten die je maar bleef aanvoeren. Mede door jouw inzet had ik genoeg patiënten om associatie analyse te kunnen verrichten, wat toch de helft van dit boekje vormt! Jos Meijer, jij bedankt voor het reviseren van alle biopten. Ik denk dat we daardoor de best gekarakteriseerde populatie van coeliakiepatiënten in de wereld hebben. Bart Crusius en Salvador Peña, bedankt voor de samenwerking bij verschillende studies en het bijdragen van patiënten en families.

Mijn vrienden wil ik bedanken voor de broodnodige ontspanning en gezelligheid. En Rob, thank you so much voor de spreadsheet die je voor me gemaakt hebt. Zonder dat was ik waarschijnlijk gek geworden! Leuk dat je mijn paranimf wilt zijn.

Nel & Karel, jullie bedankt voor al jullie knuffels en de gezelligheid tijdens het thuiswerken. Alleen jammer dat we altijd om de muis moesten vechten!

Pa & ma, bedankt voor alle mogelijkheden die jullie me geboden hebben en voor de steun tijdens de afgelopen jaren.

En tot slot, Scoob (of Rob, anders weet de helft niet over wie ik het heb!), bedankt voor alle steun die je mij hebt gegeven als het weer eens tegen zat en voor het delen van de vreugde als er een succesje geboekt was. Ik ben erg blij dat je altijd met me meeging naar de patiëntendagen, ook al werd je bedolven onder vragen van mensen die dachten dat jij ook van "het coeliakieproject" was.

Curriculum Vitae

Martine van Belzen werd geboren op 15 maart 1973 te Barendrecht. Na het eindexamen atheneum aan het Montessori lyceum te Rotterdam begon zij in 1991 aan de studie scheikunde aan de Katholieke Universiteit Nijmegen (KUN). Haar hoofdvakstage vond plaats bij de vakgroep Chemische Microbiologie aan de KUN, onder leiding van prof. dr. Vogels. Het doel van deze stage was het karakteriseren van een onbekende methaanbacterie. Vervolgens deed zij een uitgebreide bijvakstage Klinische Chemie bij het laboratorium voor Kindergeneeskunde en Neurologie van het St. Radboud ziekenhuis in Nijmegen. Deze stage werd uitgevoerd onder leiding van dr. van den Heuvel en prof. dr. Trijbels en had als doel mutatie-analyse te verrichten van het *ATM* gen bij patiënten met de zeldzame ziekte ataxia telangiectasia. In april 1998 studeerde zij af, waarna zij in juli 1998 begon aan haar promotie-onderzoek bij de afdeling Medische Genetica van het Universitair Medisch Centrum Utrecht, wat uiteindelijk heeft geleid tot dit proefschrift.

Publications

Van Belzen MJ, Meijer JWR, Sandkuijl LA, Bardoel AFJ, Mulder CJJ, Pearson PL, Houwen RHJ and Wijmenga C. A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* 2003; **125**: 1032-1041.

Van Belzen MJ, Vrolijk MM, Meijer JWR, Crusius JBA, Pearson PL, Sandkuijl LA, Houwen RHJ and Wijmenga C. A genomewide screen in a four-generation Dutch family with coeliac disease: evidence for linkage to chromosomes 6 and 9. *Am J Gastroenterol*, in press.

Seegers D, Borm MEA, van Belzen MJ, Mulder CJJ, Bailing J, Crusius JBA, Meijer JWR, Wijmenga C, Peña AS and Bouma G. Lack of association between IL12B and IRF1 gene polymorphisms and susceptibility to celiac disease. *Eur J Immunogenet*, in press.

Van Belzen MJ, Mulder CJJ, Pearson PL, Houwen RHJ and Wijmenga C. The tissue transglutaminase gene is not a primary factor predisposing to celiac disease. *Am J Gastroenterol* 2001; **96**: 3337-3340.

Van Belzen MJ, Hiel JA, Weemaes CM, Gabreels FJ, van Engelen BG, Smeets DF and van den Heuvel LP. A double missense mutation in the ATM gene of a Dutch family with ataxia telangiectasia. *Hum Genet* 1998; **102**: 187-191.

List of Abbreviations

AFBAC	affected family-based control
AGA	antigliadin IgA
AGG	antigliadin IgG
APC	antigen presenting cell
ASP	affected sibpair
CETDT	conditional extended TDT
CTLA4	cytotoxic T lymphocyte-associated protein 4
df	degrees of freedom
DZ	dizygous
EMA	anti-endomysium IgA
ESPGAN	European Society for Paediatric Gastroenterology and Nutrition
EST	expressed sequence tag
GFD	gluten-free diet
HLA	human leukocyte antigen
IBD	identical by descent
ID	individual
IFN	interferon
IL	interleukin
LD	linkage disequilibrium
MHC	major histocompatibility complex
MMLS	multipoint maximum lod score
MZ	monozygous
NCV	Nederlandse Coeliakie Vereniging
NPL	non-parametric lod score
OR	odds ratio
PCR	polymerase chain reaction
RFLP	restriction fragment length polymorphism
RR	relative risk
SNP	single nucleotide polymorphism

TDT	transmission disequilibrium test
TGM2	tissue transglutaminase gene
TNF	tumour necrosis factor
tTG	tissue transglutaminase