Unraveling the molecular genetic aspects of intestinal inflammatory disorders

ISBN 978-90-9021857-1

The research described in this thesis was carried out in the Complex Genetics Group at the Department of Biomedical Genetics of the University Medical Center Utrecht (UMCU), and within the Utrecht Graduate School for Developmental Biology. The project was financially supported by grants from the Netherlands Organization for Scientific research (grant 912-02-028), and the Celiac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch Government (grant BSIK03009).

Financial support for the publication of this thesis was provided by the Complex Genetics Group at the Department of Biomedical Genetics of the University Medial Center Utrecht, the University of Utrecht, the Section Experimental Gastroenterology of the Netherlands Society of Gastroenterology (NVGE), the Nederlandse Coeliakie Vereniging, Genome Diagnostics B.V. and the Celiac Disease Consortium.

The cover shows the silhouette of my son Allard, who tastes for the first time solid food. This was a new experience for him and I just hoped he liked it. Some parents who know that their child is at risk of celiac disease may at that moment not wonder if their child likes it, but if their child's intestine will accept the gluten in the food or react against it.
The back shows an overview figure of the research presented in this thesis, see page 156, and a free translation of a poem by J. Monsuur.

Unraveling the molecular genetic aspects of intestinal inflammatory disorders

Het ontrafelen van de moleculair genetische aspecten van ontstekingsziekten in de darm
(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
3 juli 2007 des middags te 2:30 uur

door
Alida Jentien Wijmenga-Monsuur
geboren op 23 maart 1978 te Eemnes

Promotor: prof.dr. C. Wijmenga

Contents

Contents

Preface and outline of the thesis

Preface

Celiac disease (CD) is characterized by a chronic immune reaction in the small intestine to the gluten proteins that are present in a (Western) daily diet, derived from wheat, barley and rye. It has a prevalence of 1% in the Western population (1) and is a complex disorder in that it involves both environmental and genetic factors. There is still little general awareness of the disorder or of what it involves, despite the fact that CD was observed in patients around 250 A.D. and the involvement of wheat gluten was identified in 1952 by the Dutch pediatrician Willem Karel Dicke.

For a long time CD was considered to be a pediatric intestinal disorder, but it is now recognized as a life-long disorder that can start causing symptoms at any age. Besides the intestinal problems (chronic diarrhea, growth retardation, weight loss, abdominal pain, vomiting, bloating, distention and constipation), more systemic symptoms are also being recognized (bone problems, ataxia, reproductive problems, and skin manifestations (dermatitis herpetiformis)) (2). One important complication in a small proportion of untreated CD patients is the development of enteropathy-associated T-cell lymphoma (EATL), which has an extremely poor prognosis (3). It has been noted that not all patients show symptoms (4). The lack of awareness of the occurrence of CD in adults, the presence of systemic symptoms, and the silent cases result in patients being misdiagnosed and/or not being diagnosed correctly for many years. It is estimated that only one out of seven patients (14%) is correctly diagnosed.

CD results from the uptake of gluten leading to an immune response in genetically susceptible individuals. The gluten proteins are difficult to digest in the intestine, leaving long proteins that can be toxic. Individuals that carry certain variants of the *HLA-DQA1* and *-DQB1* genes (those that form the DQ2.5 and DQ8 molecules) are able to present gluten to the T cells of the immune system, which then responds by attacking the surface of the intestine. The surface area of the small intestine normally covers a whole football field, due to the folded structure of its crypts and villi. The attack on the immune system results in a flattening of the intestinal barrier (smaller and less specialized surface), a diminished uptake of nutrients, and less protection against pathogens, leading to the symptoms described above. The only known treatment is the life-long exclusion of gluten from the diet, which allows the intestine to recover and reverse the damage.

CD is diagnosed by the presence of antibodies in the blood of patients and by observing that the intestinal barrier is flat. This observation, seen as the golden standard in diagnosing CD, is done by taking an intestinal biopsy from potential CD patients. The flattening of the intestinal barrier has been classified by the pathologist M.N.

Marsh into several stages (Marsh 0 to Marsh III) to determine the severity of the flattening (5, 6).

We still know little about this complex genetic disorder, although we do know that the main environmental factor is gluten and that there are two genes involved that together form molecules that allow gluten to be presented to the T cells. Almost all patients (over 95%) carry one of the variants of the HLA genes involved in CD, but more than 30% of the general population also carries these variants without developing CD. This, together with the prevalence of 1% in the general population, the increased prevalence of 10% in first-degree family members of CD patients, and a concordance rate of 85% in monozygotic (genetically identical) twins suggests that CD is a disorder involving multiple susceptibility genes (1, 7, 8). CD is thus considered to be a complex genetic disorder (multiple genes as well as environmental factors involved). It is estimated that the HLA genes contribute 40% of the genes involved in CD, leaving another 60% of the genetics of CD unexplained. It is also unknown which other molecular pathways, besides the immune response, play a role. Work in the past decade had resulted in the localization of regions on the chromosomes that harbor genes involved in CD (e.g. Van Belzen et al., 2003 and Babron et al., 2003) (9, 10), but had not yet led to identifying these genes. There may also be other, undiscovered, chromosomal areas that could harbor more susceptibility genes.

The aim of this thesis was therefore to find some of the genes underlying the susceptibility to CD. We hoped these would lead to a better understanding of the mechanisms involved in CD and reveal more about the complex genetic background of CD and other related disorders.

Outline of the thesis

This thesis is divided into five parts and starts with a general introduction to celiac disease and *MYO9B*, the gene we found to be associated with it (**part 1**). The thesis then follows the line that is also used in the introductory review: the entry of gluten into the intestine (**part 2**), gluten's passage through the intestinal barrier (**part 3**), and its presence in the body where an immune reaction can be evoked against gluten (**part 4**). **Part 5** is a discussion on complex genetic disorders, *MYO9B*, and the knowledge we gained on the genetics of CD.

Part 1 gives an overview on the current status of CD and *MYO9B*, the gene we found to be associated to CD.

Chapter 1.1 gives an overview of the current status of CD; it covers the general

aspects of CD as well as current insight into important molecular aspects. The role of susceptibility genes in CD is evaluated by following gluten along its path from ingestion to uptake in the body, which leads us through the three aspects of CD pathology. The first is the presence of gluten in the lumen of the intestine, where it is broken down by several enzymes. The second is the intestinal barrier through which gluten peptides pass. The third is the reaction of the immune system in response to gluten peptides, in which both the innate and the adaptive immune systems play a role. These three aspects are used as the story line throughout this thesis (see **parts 2, 3 and 4**).

Chapter 1.2 gives some background on the gene *myosin IXB* (*MYO9B*) that we found to be associated to CD. We hypothesize about the possible roles of *MYO9B* in CD and other immune-related disorders.

In **Part 2** we look into two enzymes that are involved in gluten breakdown in the lumen of the intestine

Chapter 2.1 describes the genetic and functional analysis of *pyroglutamyl-peptidase I* (*PGPEP1*) in CD. This gene is located in the Dutch CD linkage region on chromosome 19p13.1, and is able to cleave the undigested parts of gluten in the intestine. It may thus play a role in the susceptibility of CD patients to gluten.

Chapter 2.2 describes the genetic and functional analysis of *prolyl endopeptidase* (*PREP*) in CD. This gene is located in the Dutch linkage region on chromosome 6q21-22 and is the only gene able to directly cleave gluten. Bacterial forms of this gene are now being tested for their possibility to cleave gluten into harmless peptides before they enter the intestine.

In **Part 3** we describe our search for genes involved in the increased permeability of the intestinal barrier in CD

Chapter 3.1 describes the discovery of *MYO9B* as the susceptibility gene in the Dutch CD linkage region on chromosome 19p13.1. The possible involvement of *MYO9B* in the increased permeability of the intestinal barrier in CD and the observation that this increased permeability plays a role in multiple immune-related disorders led the research focus for the following chapters.

In the studies described in **chapters 3.2 and 3.3** we tried to replicate our findings in other populations. Replication is important in complex genetic studies to exclude the possibility of statistical type I errors, but it is often hampered by the low relative risk of these genes and the resulting low power of most studies, as well as the fact that the true causal variant in the gene may not have been discovered.

In **chapter 3.4** we report the possible involvement of *MYO9B* in two other im-

mune-related disorders – Crohn's disease and ulcerative colitis, which are grouped under the name inflammatory bowel disorders (IBD). These two disorders also show linkage to chromosome 19p, and share the feature of increased intestinal permeability observed in CD.

Chapter 3.5 describes our search for other genes involved in the increased intestinal permeability by studying the tight junction genes. Tight junctions close the gaps between the enterocytes, a cell layer that lines the intestine, resulting in a physical barrier between the body and the outside world and giving protection against unwanted antigens or pathogens and allowing a selective uptake of nutrients. In this study we used our CD cohort for the discovery of the associated genes, as well as a British cohort to replicate our results. We also looked into the role of the associated genes in IBD.

In **Part 4** we describe searches for genes involved in the inflammatory component of CD and looking for a good method to screen for the HLA risk factors

Chapter 4.1 describes the search for associated genes involved in inflammation. We looked for these genes in CD in two linkage regions, on chromosomes 19p13.1 and 5q31. This was based on our hypothesis that CD and IBD could share genetic factors. They show overlap in their pathogenesis (e.g. inflammatory responses) as well as in linkage regions, which could indicate that these regions contain genes involved in the susceptibility to both disorders.

Chapter 4.2 describes looking for a good method to screen for the risk genes *HLA-DQA1* and *-DQB1*. Variants of these genes are capable of presenting gluten to the T-cells of the immune system, thereby empowering an immune response that leads to the damage in the intestine, which is seen as the main problem in CD. Over 95% of CD patients carry these risk variants. Better and more simple screening methods, like the one described in this chapter, are needed to get a step closer towards population screening for CD, which is necessary to find all the undiagnosed patients who are at risk of possibly severe complications.

Part 5 discusses the knowledge gained from the work for this thesis

The discussion focuses on three parts: the complex search for genes involved in complex genetic disorders, the role of MYO9B in CD, and the knowledge gained from our genetics studies in CD. It ends with some future perspectives on the research for CD and other complex genetic disorders in general.

References

1. Dube C, Rostom A, Sy R, Cranney A, Saloojee N, Garritty C, et al. The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. Gastroenterology. 2005 Apr;128(4 Suppl 1):S57-67.
2. Rewers M. Epidemiology of celiac disease: what are the prevalence, incidence, and progression of celiac disease? Gastroenterology. 2005 Apr;128(4 Suppl 1):S47-51.
3. Daum S, Cellier C, Mulder CJJ. Refractory coeliac disease. Best Practice & Research Clinical Gastroenterology. 2005;19(3):413-24.
4. Dewar DH, Ciclitira PJ. Clinical features and diagnosis of celiac disease. Gastroenterology. 2005;128(4):S19-S24.
5. Marsh M. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity (‘celiac sprue’). Gastroenterology. 1992;102(1):330-54.
6. 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 Apr;94(4):888-94.
7. Greco L, Romino R, Coto I, Di Cosmo N, Percopo S, Maglio M, et al. The first large population based twin study of coeliac disease. Gut. 2002 May 1, 2002;50(5):624-8.
8. Nistico L, Fagnani C, Coto I, Percopo S, Cotichini R, Limongelli MG, et al. Concordance, disease progression, and heritability of coeliac disease in Italian twins. Gut. 2006;55:803-8.
9. Babron MC, Nilsson S, Adamovic S, Naluai AT, Wahlstrom J, Ascher H, et al. Meta and pooled analysis of European coeliac disease data. Eur J Hum Genet. 2003 Nov;11(11):828-34.
10. Van Belzen MJ, Meijer JW, Sandkuijl LA, Bardoel AF, Mulder CJ, Pearson PL, et al. A major non-HLA locus in celiac disease maps to chromosome 19. Gastroenterology. 2003 Oct;125(4):1032-41.

Part 1

Overview on the current status of celiac disease and myosin IXB

Chapter 1.1

Understanding the molecular basis of celiac disease: What genetic studies reveal

Alienke J. Monsuur & Cisca Wijmenga

Annals of Medicine, Volume 38, Issue 8 December 2006, pages 578-591

TRENDS IN MOLECULAR MEDICINE

Understanding the molecular basis of celiac disease: What genetic studies reveal

ALIENKE J. MONSUUR & CISCA WIJMENGA

Complex Genetics Section, Department of Biomedical Genetics, UMC Utrecht, The Netherlands

Abstract
Celiac disease (CD) is characterized by a chronic immune reaction in the small intestine to the gluten proteins that are present in a (Western) daily diet. Besides the well known involvement of the HLA class II histocompatibility antigen (HLA)-DQ2.5 and -DQ8 heterodimers (encoded by particular combinations of the HLA-DQA1 and -DQB1 gene) in CD and the minor contribution of the *CTLA-4* gene, recently the myosin IXB (*MYO9B*) gene has also been found to be genetically associated. This review covers the general aspects of CD as well as current insight into important molecular aspects. We evaluate the role of susceptibility genes in CD by following gluten along its path from ingestion to uptake in the body, which leads us through the three aspects of CD's pathology. The first is the presence of gluten in the lumen of the intestine, where it is broken down by several enzymes. The second is the intestinal barrier through which gluten peptides pass. The third is the reaction of the immune system in response to gluten peptides, in which both the innate and the adaptive immune systems play a role. Our main conclusion, based on the current genetic and functional studies, is that we should look for causal genes in the barrier function as well as in the immune systems.

Key words: *Adaptive immunity, Celiac disease, CTLA-4, gluten, HLA-DQ, intestinal barrier, MYO9B, myosin IXB, innate immunity, tissue transglutaminase*

Introduction

An inflammatory disorder in response to gluten

Celiac disease (CD) is characterized by a chronic immune reaction in the small intestine to the gluten proteins that are present in a (Western) daily diet. Gluten proteins are storage proteins present in wheat, barley and rye, and are needed to maintain the processing quality of food derivates such as bread, pasta, and cookies; they are also widely used to bind and thicken sauces. CD mostly occurs in Westernized populations where gluten-derived products form a large part of the diet. The involvement of the *HLA*-DQ2.5 and -DQ8 heterodimers (encoded by particular combinations of the HLA-DQA1 and -DQB1 gene) in CD is well known since most of the patients carry these molecules.

This review covers the current insight into important aspects of CD and evaluates the role of susceptibility genes in CD by following the gluten molecules from the lumen of the small intestine, passing through the epithelial barrier of the intestine and eliciting the immune response (see also Figure 1).

General overview of CD

Clinical features and treatment

Until recently CD was considered to be a pediatric intestinal disorder, seen mostly in young children who had just started a gluten-containing diet.

The intestinal problems seen in CD patients comprise chronic diarrhea, growth retardation, weight loss, abdominal pain, vomiting, bloating, distention, and constipation (reviewed in Rewers et al.) (1). CD is now being recognized more and more in older patients who do not present with all the clinical features or who are developing more of these features at a later age. Their symptoms are more systemic, and the disease is now more likely to present as a multiorgan disorder, including bone problems, ataxia, carditis, reproductive problems and skin manifestations like dermatitis herpetiformis

Correspondence: Professor Cisca Wijmenga, Complex Genetics Section, Stratenum 2.117, Department of Biomedical Genetics, University Medical Centre Utrecht, PO Box 85060, 3508 AB Utrecht, The Netherlands. Fax: +31-30 253 8479. E-mail: t.n.wijmenga@med.uu.nl

Abbreviations

ASCA	anti-Saccharomyces cerevisiae antibodies
CARD	caspase recruitment domain
CD	celiac disease
DH	dermatitis herpetiformis
RCD	refractory CD
EATL	enteropathy-associated T cell lymphoma
RR	relative risk
MYO9B	myosin IXB
CTLA-4	cytotoxic T lymphocyte-associated 4
PREP	prolyl endopeptidase
PGPEPI	pyroglutamyl-peptidase I
tTG	tissue transglutaminase
IBD	inflammatory bowel disorders
MYLK alias *MLCK*	myosin light chain kinase
ROCK	Rho kinase
TNF	tumor necrosis factor
MICA	MHC class I polypeptide-related chain A
IEL	intraepithelial lymphocytes

(2). These older patients may present with either active or silent CD (reviewed in Dewar et al.) (2). The more systemic presentation of active CD leads to these patients not always being recognized or properly diagnosed. The silent cases are unaware of their disease due to a lack of symptoms, or they may have only vague symptoms which often remain undiagnosed. They are mostly identified during the screening of large populations to determine prevalence, or of other risk groups like index patients' family members.

Over the last 25 years it was observed that the frequency of patients presenting with diarrhea at diagnosis has gradually dropped from 91% to 37% (3). The duration of symptoms before diagnosis has also decreased, from 11 years to 4 years, nonetheless this is just the average duration and 4 years is still a long period, meaning that there are many patients who have remained undiagnosed for a long time.

The only existing treatment for CD is a life-long gluten-free diet (GFD), which reverses the damage in the intestine so that the Marsh III stage (Marsh stages, see diagnosis and Table I) gradually changes to a Marsh 0 stage, the clinical symptoms disappear, and the risk of complications diminishes. The normalization in the intestine is slower in adults than in young children, who often reach the Marsh 0 stage in a few months, while adults may take up to a few years (4).

Key messages

- Gluten proteins are resistant to enzymatic breakdown, leading to peptides that may evoke an immune response in genetically susceptible individuals.
- The intestinal barrier is impaired in celiac disease patients and this affects its protective function against pathogens and the selective uptake of nutrients. Myosin IXB was found to be genetically involved in the impaired barrier.
- Both the innate and the adaptive immune response play a role in celiac disease and the genetic involvement of the HLA-DQ2.5 and -DQ8 molecules as well as the *CTLA-4* gene has been shown.

Features observed in CD patients are an impaired intestinal barrier, less regular tight junctions and altered sugar absorption ratios. These features are mostly seen in patients on a gluten-containing diet, although the functionality of the intestinal barrier does not recover completely in patients on a gluten-free diet who regain a normalized intestine (5–7). Patients with dermatitis herpetiformis (DH, discussed below) who show no histological damage in the intestine may also have an impaired intestinal barrier (8).

One of the systemic manifestations of CD is DH generally seen as a cutaneous manifestation of CD (reviewed in Zone et al.) (9). Most DH patients have varying degrees of lesions in the small intestine and react to gluten. DH and CD run in families and often co-occur in individuals. The best treatment is a gluten-free diet, although most patients only treat their skin problems using medicine and neglect the underlying intestinal problems caused by gluten.

A small proportion (<7%) of CD patients do not respond to a strict GFD (4). This group of CD patients (generally referred to as refractory CD (RCD) patients) is split into RCD type I and RCD type II patients and was defined by Daum et al. as individuals with a persisting villous atrophy with crypt hyperplasia and increased intraepithelial lymphocytes (IEL) in spite of being on a strict GFD for more than 12 months, or those whose severe and persisting symptoms necessitate intervention independent of the duration of their GFD (10). The RCD type II group (<0.3% of the total CD population) shows abnormal IELs, with abnormal surface markers in small intestinal biopsies, like the expression of intracytoplasmic CD3e and the lack of several classical T cell markers, e.g. CD8 and CD4.

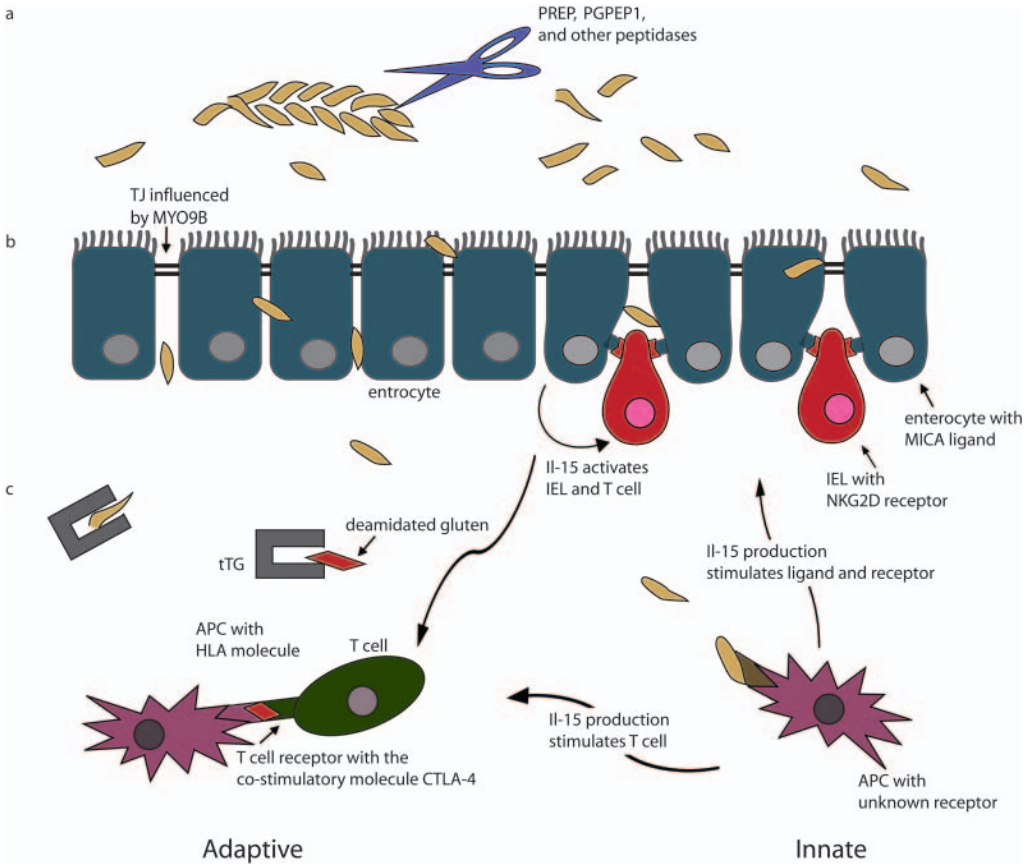


Figure 1. Overview of the path taken by gluten. a) Gluten enters the lumen of the small intestine and is cleaved by several enzymes. b) It subsequently passes through the enterocytes in the intestinal barrier, which is impaired in celiac disease (CD) patients; myosin IXB (*MYO9B*) might play a role here by affecting the tight junctions. Enterocytes express the MICA ligand that binds to the NKG2D receptor on the intraepithelial lymphocytes (IEL). Besides this, enterocytes also produce IL-15 capable of stimulating IELs and T cells. c) Both the innate and the adaptive immune system play a role. 1) Adaptive immune response: gluten is deamidated by tissue transglutaminase (tTG) in the lamina propria and elicits an immune response in genetically susceptible individuals who are also HLA-DQ2.5 and/or -DQ8 positive. 2) Innate immune response: gluten induces the antigen-presenting cell (APC) to produce Il-15, which stimulates the production and interaction of the MICA ligand and the NKG2D receptor. IL-15 also stimulates the adaptive immune response.

Table I.

Marsh stage	Intraepithelial lymphocytes (IEL) >30 per 100 epithelial cells	Crypt hyperplasia	Villous atrophy [a]
0 [b]	–	–	–
I	+	–	–
II	+	+	–
IIIA	+	+	partial
IIIB	+	+	subtotal
IIIC	+	+	total

[a] Broadening and disappearing villi, with a crypt/villous ratio >1.
[b] Biopsies taken from celiac disease (CD) patients who are on a gluten-free diet and who returned to a normalized histology. The histology should basically be the same as in healthy controls.

A proportion of this group develops enteropathy-associated T cell lymphoma (EATL).

CD also co-occurs with other autoimmune disorders in families and in patients (11), suggesting an overlap in causative genes between different disorders.

Diagnosis

Serological screening is often used as the first step in the diagnosis of CD, measuring the endomysial antibodies and tissue transglutaminase antibodies; the latter is the autoantibody in CD (see review by Schuppan et al.) (12). A diagnosis of CD is made

when a person's symptoms meet the revised ESPGHAN criteria (13). The main diagnostic criteria are a positive biopsy (Marsh III A, B, or C) while on a gluten-containing diet, and a full clinical remission when the patient is on a gluten-free diet. Duodenal biopsies are classified according to Marsh who defined four stages, the last of which was later divided into three substages (see Table I) (14,15). Positive antibodies that disappear when the patient is on a gluten-free diet add to the diagnosis.

If there is any doubt about the initial diagnosis, extra steps should be taken. A doubtful diagnosis may occur more often in children under the age of 2 years, who may also suffer from other causes of enteropathy (like giardiasis, cows' milk sensitive enteropathy, and postenteritis syndrome). A biopsy on a gluten-free diet should be taken—this should show a normalized histology—and then they should be given a gluten challenge after 2 years (or longer) on a gluten-free diet. The gluten challenge should not be given before the age of 6, and it needs to be followed by a biopsy, which should again show a Marsh III classification (13).

Prevalence and genetics

The prevalence of CD in the general population ranges from 0.5% to 1.26% and is estimated to be on average 1%, based on serology screenings in unselected populations (systematically reviewed by Dube et al.) (16). These numbers have also been found in the USA, where CD was long thought to be a rare disorder (17,18). The prevalence in patients' family members ranges from 2.8% to 22.5% depending on the study methods (i.e. serology and/or biopsy) and on the groups tested (e.g. first/second degree family members, two affected patients already known in the family) (16). This implies an increased familial clustering of ~10%, resulting in a relative risk (RR) of 10 for the disorder and suggesting a role for genetic factors. The prevalence was further corroborated by two large population-based twin studies in the Italian population, which found a concordance rate in dizygotic twins of around 20% and in monozygotic twins of around 85% (19,20). The increase in concordance between twins and the fact that monozygotic twins do not have a 100% concordance rate suggests the involvement of environmental factors in addition to genetic factors.

A complex genetic disorder

CD is a complex genetic disorder involving multiple genes as well as environmental factors, of which the most important is gluten. For the genetic part of CD we expect to find multiple genes since inheritance of CD does not follow a Mendelian pattern. We assume a model comprising a major gene (the HLA gene that will be discussed below) and several low-risk genes (such as myosin IXB (MYO9B) and cytotoxic T lymphocyte-associated 4 (CTLA-4)) (21–26). Discovering the gene(s) in complex disorders is a daunting task and past progress has been slow, but due to modern techniques more and more causative genes will now be found (Box 1; Table II). The loci presented in Table II are reviewed in detail by van Heel et al. (27).

HLA gene has been the only known genetic variant for over 30 years

The involvement of the HLA complex located in the major histocompatibility complex (MHC) region on chromosome 6, which predisposes to CD, has been known for over 30 years (28–30). Via the observed involvement of HLA-A8, HLA-DW3 (now HLA-DR3) and HLA-DQ in CD (28,29,31–33), the

Table II.

Locus name	Location	λs [e]	Expected role in CD	Gene
CELIAC1	6p21.3	4,6	major	*HLA-DQA1* and *-DQB1*
CELIAC2	5q31–q33	ND	minor	
CELIAC3	2q33	ND	minor	*CTLA-4*
MYO9B, previously CELIAC4	19p13.1	2,6	intermediate	*MYO9B* [d]
ND	15q11–q13	ND	minor [a]	
ND	9p21–p13	ND	minor [b]	
ND	6q21–22	2,3	minor [c]	

[a] This locus was found using large pedigrees in which a different inheritance pattern is observed than in the general population.
[b] This locus was found in a large pedigree with a dominant inheritance, and in two other studies it showed some evidence for linkage.
[c] This locus was only found once in a celiac disease (CD) population, with suggestive evidence, but is also found in several other immune-related disorders.
[d] Relative risk of this gene, allele frequency=1.5, heterozygous AG=1.66 and homozygous AA=2.27.
[e] λs as calculated by Van Belzen et al. (104).
ND=not determined

HLA-DQA1 and -DQB1 genes were shown to be involved and were the first genes found to be genetically and repeatedly associated to CD (34–37). The HLA-DQA1 and -DQB1 genes are in a region of high linkage disequilibrium and together form heterodimers, some forms of which are associated to CD. The estimated contribution of the HLA region on developing CD is around 40% (38,39), meaning the remaining 60% of genetic factors involved in CD was unknown.

Recently a second gene, *MYO9B* on chromosome 19, was identified in the Dutch population (40); it may explain some 20% of the genetic risk factors involved in the Dutch CD population. Replication in other populations is needed to determine the true relative risk. At the moment there is no positive replication, which might be due to lack of power and overestimation of the genetic risk (a problem that is often seen in genetic studies; see Box 1 for a more extensive discussion of genetics studies, with MYO9B as an example) (41–43). Besides this, the *CTLA-4* gene that plays a role in several immune disorders is also involved in CD, although it probably only accounts for a few percent of the genetic variation (21–26). Since we do not know how these three genes interact, it is difficult to determine the exact genetic space left for other possible genes.

Molecular pathogenesis of CD

The molecular aspects of CD can be divided into three parts; these will be elaborated in the rest of this review.

The main environmental factor involved in CD is gluten. Following its path from ingestion to uptake in the body leads us through three aspects of the pathology of CD. The first is the presence of gluten in the lumen of the intestine, where it is broken down by several enzymes into smaller peptides. The second is the intestinal barrier through which gluten peptides pass; it is the divide between the body and the external world. The third is the reaction of the immune system in response to gluten peptides. The genes that have been implicated in CD (Table II), and the ones still to be found, may be involved in all these aspects and reveal more about the molecular basis of CD.

Gluten in the lumen of the small intestine

Gluten consists of multiple immunogenic peptides

Gluten is a mixture of gliadin and glutenin proteins, which, in turn, are mixtures of α-, β-, γ-, ω-gliadins, low molecular weight glutenins, and high molecular weight glutenins, each having allelic variants resulting in a huge number of functional and nonfunctional proteins (44). A subset of these glutens, the α- and γ-gliadins, low molecular weight glutenins, and high molecular weight glutenins, can give rise to potential toxicity. Gluten proteins and their homologues are present in different grains: wheat, rye and barley. Pasta wheat consists of a combination of two genomes giving rise to the AABB tetraploids, while bread wheat consists of a combination of three genomes, giving rise to the AABBDD hexaploids. Especially the D genome contains the toxic gluten epitopes. Some cultivars of the pasta wheat (tetraploid AABB genome) contain no toxic gluten epitopes. This implies that it is possible to find or modify cultivars that lack the toxic epitopes, have baking qualities and might be safe for CD patients. Several of the existing wheat varieties have been tested for their potential toxicity, and they show differences in toxic T cell responses (45–47). One example of a naturally occurring, nontoxic wheat variety is Teff (48).

Several enzymes play a role in gluten break down

Dietary proteins are broken down by pepsin in the stomach, secreted pancreatic proteases and carboxypeptidases. Then exo- and endopeptidases located in the brush border membrane of the epithelial layer of the small intestine continue the digestion into mono-, di- and tripeptides that can then be transported across the epithelial cells into the lamina propria.

Unlike most dietary proteins, gluten is indigestible by most proteases due to its high proline rich content (49). The only gluten-specific enzyme is prolyl endopeptidase (PREP), which cleaves peptide bonds at the C-terminal side of proline residues (50). Other peptides can play a role in the subsequent digestion steps, like pyroglutamyl-peptidase I (PGPEPI), which has been shown *in vitro* to cleave the indigestible parts of gluten, probably when L-pyroglutamyl residues are being formed after the first cleavage steps (51). This cleavage leads to unstable peptides that are easily digestible.

It has been hypothesized that small aberrations in the function of one of the proteases could lead to longer peptides or an increased load of toxic gluten epitopes in the lumen or lamina propria (52). A decrease in gluten digestion by brush border enzymes has been observed in untreated CD patients when compared to controls (53). This resistance of gluten to enzymatic digestion might result from the changes in cell function of the intestine secondary to other pathological mechanisms, or it could be causal

due to a genetic background that makes patients more vulnerable to gluten.

Following this, the two genes that encode for peptidase enzymes, *PREP* and *PGPEP1*, are also located in CD linkage regions (6q21–22 and MYO9B, respectively) and have been studied for their potential causal role in CD (Table II), but no association could be found (54,55).

This suggests that if these enzymes are indeed involved in the less proper digestion of gluten, their role in CD is not causal but secondary, and it is rather the impaired structure of the intestine that influences the function of these enzymes.

Possible treatment with a bacterial form of PREP

The 33-mer gliadin peptide that cannot be digested by gastric and pancreatic enzymes was shown to be digestible using a bacterial homologue (from *Flavobacterium meningosepticum*) of prolyl endopeptidase called PEP (52). Peptidase therapy might be a potential therapy, since this digestion resulted in a decrease of immunogenic gluten peptides, thereby diminishing the T cell response and the specificity to tissue transglutaminase (tTG, discussed below) (56).

The epithelial layer

Forms the barrier between the body and the external world

The body tissue is separated from bacterial-filled lumen of the intestine by a single layer of cells, the enterocytes. The enterocytes have two major functions: to form a physical barrier as protection against unwanted antigens/pathogens (like microorganisms), and the selective uptake of nutrients. This important intestinal layer is more permeable in CD patients and some of their relatives, since tight junctions are less regular and sugar absorption ratios are altered (5–8). Normally this layer has an enormous surface area (equivalent to that of a soccer field) due to the folds in the villi and microvilli. This surface area is drastically reduced in CD patients. The stem cells in the crypt give rise to progenitor cells that move up along the crypt/villus axis, they lose their capacity to proliferate, and mature into differentiated enterocytes. As these cells after approximately 5 days become too old, unnecessary, or infected, they are shed into the lumen by an unknown but well orchestrated process that does not seem to involve apoptosis as a major or triggering factor. Interestingly, in spite of the constant cell shedding from this single epithelial cell layer, a study in mice has shown that its integrity does not seem to be impaired although gaps in this single cell layer can be seen for up to 1 hour (57). Tight junctions between the cells are one of the mechanisms to retain the barrier integrity, but there are also some undefined fluids seen in mice that fill the gaps until they are resolved.

Besides being a physical barrier, the enterocytes have several other functions, like uptake of small peptides, and defense against and cross-talk with pathogens. The vast amount of bacteria and other microorganisms in the lumen of the intestine that can either be beneficial or pathogenic means there is a need for cross-talk between them and the enterocytes (58). This cross-talk is done partly by pattern recognition proteins (encoded for by Toll-like receptors and caspase recruitment domain family genes (CARD)) leading to either a response against the pathogen or a tolerance to it (for more insight into the intestinal epithelial barrier see Ismail et al.) (58).

One of the peptides that have to pass the epithelial barrier is gluten. We do not know whether gluten peptides cross this barrier by a paracellular or transcellular mechanism, making it difficult to assess how the impaired barrier increases susceptibility to CD.

An intriguing finding in CD patients is that the normally almost sterile proximal small intestine contains a high number of rod-shaped bacteria on the mucosa (59). This feature is also seen in treated patients and seems to indicate that bacterial penetration and subsequent binding is promoted in CD patients. This might be due to an altered mucous/glycocalyx layer, although the origin of the process is unknown. Perhaps the increased stickiness of the epithelial layer to prolamin may irritate the epithelial surface, and prolamin is then mistaken for a pathogen giving rise to the immune system's unwanted reaction seen in CD patients.

Increased permeability of the epithelial barrier is seen in CD patients

CD patients show an impairment of the epithelial barrier but exactly how the damage is caused cannot be pinpointed to one mechanism. It was long thought that the immune response to gluten causes a self-sustaining cycle of tissue damage and the release of tissue transglutaminase (tTG, discussed below) with subsequent repair, but enormous cell damage is not observed in the tissue and the reality is probably more complicated. The impairment could be due to an altered ratio of proliferation and differentiation of the enterocytes (60). The enterocytes might therefore lose some of their specialized functions. Another mechanism could be

the increased permeability, due to genetic variants (5–8), seen in CD patients and some of their family members, which is greater in untreated patients but still present in treated patients and healthy relatives. The anti-*Saccharomyces cerevisiae* antibodies (ASCA) that are seen as a molecular marker of intestinal permeability are present in one-third of CD patients and partly disappear during a gluten-free diet, suggesting the barrier recovers (61). The increased amounts of interferon-γ (IFN-γ) and TNF-α seen in CD could also enhance increased permeability of the barrier (62).

Genetic association has been shown to MYO9B

Defining the exact mechanisms that make the epithelial barrier of CD patients more vulnerable will guide our search for causative factors. One of the genes involved in barrier impairment might be the recently discovered *MYO9B* gene, which is associated to CD in the Dutch population (40). Association of this gene has now also been observed with inflammatory bowel disorders (IBD) and especially with ulcerative colitis, which may share the same mechanism that is involved in barrier impairment (63). Although the function of this gene in general and its implication in CD is still undefined, the gene family and the domains provide some insight. *MYO9B* is a single-headed myosin motor gene that, due to its actin binding domain, is able to bind to the actin filaments in cells and move along them (64,65). MYO9B carries its own cargo, the Rho GTPase activation protein (Rho-GAP) domain which regulates the Rho family GTPases, to its site of action. Rho family GTPases have two functions with respect to tight junctions; they regulate the junction assembly and the selectivity of the paracellular route in the enterocytes (66). A more active RhoA (guanosine triphosphate-bound form) negatively regulates the tight junctions resulting in increased permeability. A more inactive form of RhoA (guanosine diphosphate-bound form) decreases permeability, showing that a tight regulation of RhoA is important in the balance that the intestinal border needs to exert its two main functions of being a protective and a selective barrier. A recent report connects tTG (discussed below) to RhoA activation, and genetic variants in *MYO9B* might also influence its own capability to regulate Rho family proteins and therefore influence the actin filaments, tight junctions and cell shapes resulting in the leaky barrier seen in CD patients (40,67).

MLCK and ROCK

A process that potentially had some overlap with *MYO9B* has been seen in IBD, where the gene myosin light chain kinase (*MYLK*, *MLCK*) has been shown to have a higher expression and activity in IBD patients, with the increase of expression correlating with the severity of the lesion (68). Patients with inactive disease and family members also show some increase of expression. *MYLK* is involved in myosin II activation and consequently in the functioning of the tight junctions (62). Here the proposed mechanism is either increased expression leading to a leakier barrier and reactions of the immune system, or an increase of expression due to cytokine-signaling of pro-inflammatory cytokines, like tumor necrosis factor (TNF), with the leaky barrier as a result. Both mechanisms could lead to a self-sustaining cycle of barrier impairment and inflammatory responses. In addition, *MYLK* together with Rho kinase (ROCK) is involved in purse-string wound healing (69). In the small intestine where there is constant cell shedding and pressure to maintain the barrier, an altered function of genes involved in permeability and wound closure could lead to disease, and it is perhaps here that MYO9B plays a role in the disease process, if its function has similar effects to *MYLK*.

Increased barrier is not just confined to CD: is there an overlap in disease mechanisms?

An impaired barrier function has been suggested for multiple disorders, e.g. IBD, asthma, type I diabetes and psoriasis (70–72). In all these disorders an increased reaction of the immune system is seen in response to known or unknown pathogens. This could be due to an impaired epithelial cell barrier in the intestine or in one of the other organs lined with epithelial cells (70,72–74). These disorders can also co-occur in families, which might be due to overlapping susceptibility genes, such as *MYO9B*. Although the same gene might be involved in various disorders, the causal variant may not necessarily be the same.

Gluten peptides evoke an immune response

HLA presents gluten peptides to the T cells

The first gene repeatedly shown to contribute to the genetics of CD was the *HLA-DQ* gene. The HLA-DQ2.5 molecule (built up from the DQA1*0501 and DQB1*0201 variants) predisposes to CD, because of its binding properties for gluten peptides. The HLA-DQ2.2 (built up from the DQA1*0201 and DQB1*0202 variants) predisposes to CD only when it is expressed together with the DQ2.5. This is due to the DQB1*0202 that can in combination with the DQA1*0501 of DQ2.5 form a functional

molecule with similar binding properties for gluten peptides. Lastly, the HLA-DQ8 molecule (built up from the DQA1*0301 and DQB1*0302 variants) also gives some predisposition to CD. It has been found that over 90% of CD patients carry the DQ2.5 molecule (alone or in combination with DQ2.2 or DQ8), and most of the remaining patients carry the DQ2.2 or DQ8 (75,76). Fewer than 6% of the CD patients carry none of these molecules, but some of them do carry one-half of the DQ2.5 heterodimer. In Europe a gradient is seen for the DQ types in CD patients: southern European populations have more CD patients who carry none of the risk DQ molecules by themselves, but in whom the risk molecules can form due to combinations of the HLA-DQA1 and -DQB1 genes on the different chromosomes (trans effect) than the northern European populations, in whom the combination of genes on one chromosome already forms the molecules (cis effect) (75,76). Although these molecules play a large role in CD, they cannot be the only contributing factors, since around 25% of the normal population also carry the DQ2 molecules without having CD, for example. The HLA-DQ2.5 and -DQ8 molecules are therefore seen as necessary, but not sufficient, to cause CD.

Dose-response effect

The highest RR of developing CD is seen for persons homozygous for the DQ2.5 molecule (homozygous for the variants of HLA-DQA1 and -DQB1 genes that form the DQ2.5 molecule), compared to those heterozygous for DQ2.5 or DQ8 (heterozygous for the HLA-DQA1 and -DQB1 genes that form these molecules) (76). This dose-effect has also been observed functionally (77). HLA-DQ dimers can be formed by cis- and trans-combinations. The HLA-DQA1 and HLA-DQB1 genes can therefore give rise to 0–4 functional DQ2.5 molecules, and to correspondingly increasing levels of HLA-DQ2.5 peptide complexes that can lead to immune responses when the reaction threshold of the T cells is crossed (77,78). A recent study showed that HLA-DQ2.5 homozygosity is more than doubled (from 20.7% to 44.1%) in RCD type II patients (79), and in patients with an EATL, the amount of HLA-DQ2.5 homozygosity increases to 53.3%. This suggests that HLA-DQ2.5 homozygosity increases the risk of becoming an RCD II patient or of developing an EATL.

tTG modifies gluten peptides and makes them more immunogenic

The role of the HLA-DQ complex was only partly understood, since it is not a highly potent gluten peptide binder. This changed when it was shown that tTG has an effect on the binding compatibility of gluten peptides to HLA-DQ2.5 and -DQ8, by selectively deamidating specific glutamines in gluten epitopes. This introduces negatively charged glutamic acids in the epitopes that fit better into the binding pocket of the HLA-DQ2.5 or -DQ8 molecule, and are therefore more capable of stimulating the T cells (80–83). The complex formation of tTG and gliadin in untreated CD patients is increased compared to treated CD patients and controls (84). These tTG-gliadin complexes are seen more in the epithelial and subepithelial levels and less in the lamina propria, when compared to controls. A recent paper from Sakly et al. found more tTG expressing cells in the basement membrane and lamina propria, together with an increased staining (85). The amount of tTG expressing enterocytes in cases compared to controls was decreased. It has now been shown that tTG can incorporate gliadin into the interstitial matrix components of the lamina propria, leading to an increased availability of gliadin that might act as an extra trigger for the reaction processes seen in CD (86). Although the *tTG* gene is important in the pathogenesis of CD in several ways, no causative role has been found. Firstly, there were no differences in the coding sequence—sequenced at RNA level—detected between CD patients and controls (87), and secondly a genetic study did not show any linkage or association with CD (88).

Adaptive and innate immune system in CD

For a long time it was thought that only the adaptive immunity plays a role in CD. It is now clear that both the adaptive and the innate immune responses are important in CD and that some gluten peptides seem to be involved in either one of these responses (reviewed by Jabri et al.) (89). This is also reflected by several gluten peptides having distinct pathological mechanisms in CD.

The adaptive immune response involves $CD4^+$ T cells that are activated by cells, like antigen presenting cells, which present the gluten peptides on their HLA-DQ2.5 or -DQ8 molecule. An example of a gluten peptide that stimulates the adaptive immune response via $CD4^+$ T cells of most adult CD patients is the α-gliadin 56–75 peptide (52,89–91). The priming of the $CD4^+$ T cells in CD is somewhat different from a normal $CD4^+$ T cell reaction, since interleukin-12 (IL-12) and signal transducer and activator of transcription 4 (STAT-4), which are normally active in this T cell reaction, do not seem to be involved in the process in CD. The exact role

of these $CD4^+$ T cells is still not well defined since it cannot explain all the changes in the intestine since other disorders associated with $CD4^+$ T cells do not show the increase in IELs and the malignant transformation of these IELs into EATL that occurs in some CD patients. Jabri et al. proposed a role for $CD4^+$ T cells in arming IELs (89). IFN, which is produced by the activated $\alpha\beta$ $CD4^+$ T cells, creates a more immunogenic environment and could make the enterocytes and the IEL more sensitive. This could provide one of the links between the adaptive and the innate immune responses.

Some of the genes involved in the activation of the adaptive immune response have been genetically tested for their role in CD, but no positive results were found except for the HLA-DQ2.5 and -DQ8 variants (reviewed in Diosdado et al.) (92). CTLA-4, which is a co-stimulatory molecule of the T cells, seems to have a minor involvement in CD although its role is still under debate (21–27). Generally, genes involved in the regulation of T cell activation may have a minor effect on disease susceptibility. *CTLA-4* and *PTPN22* are two genes involved in the risk to several autoimmune disorders, although especially the role of the latter in CD is not yet clear (21–27,93,94).

The role of the innate immune system in CD is becoming more and more widely recognized. The peptide that has been studied most in relation to stimulation of the innate response is the α-gliadin p31–49, which induces an immune response in the antigen-presenting cells (APCs) and epithelial cells but not in the $CD4^+$ T cells (89). This was shown by the increased production of IL-15 due to gluten-induced epithelial stress, leading to enterocytes with an upregulated MHC class I polypeptide-related chain A (MICA), a change of the cytotoxic $CD8^+$ T cells into lymphokine-activated cells and the expression of the NKG2D receptor on the IELs (95). This reduces the threshold for T cell receptor activation and mediates the direct killing of epithelial cells.

Although it has long been thought that *MICA* might play a causal role in CD, there is no genetic evidence yet. It is, however, extremely difficult to study the possible genetic effect of *MICA* in CD. *MICA* is located in the HLA region and in high linkage disequilibrium with the *HLA-DQ*, meaning that certain *MICA* variants and certain *HLA-DQ* molecules segregate together more often than would be expected by chance (96). To study the effect of *MICA* independently of *HLA-DQ*, we would need a large HLA-DQ2.5-positive case group, as well as a large HLA-DQ2.5-positive control group to see if a certain *MICA* variant is more often present in the CD patients' group compared to the control group, independently of the shared HLA-DQ2.5 background. A sufficiently large control group is more difficult to obtain since only 25% of controls carry the HLA-DQ2.5 molecule.

The IL-15 produced by APCs and enterocytes is also capable of stimulating IELs (97,98) and the T cells of the adaptive immune system, and is therefore one of the links between adaptive and innate immune system.

Discussion

To date, CD is the best understood HLA-related disorder. The HLA-DQ2.5 and -DQ8 molecules are able to present the environmental factor gluten to the T cells, especially when tTG has modified these gluten proteins in order to make them fit better into the binding pocket of the HLA-DQ molecules. The immune response that is then elicited is involved in the changes observed in the small intestine of CD patients. tTG plays several roles in CD, like modifying the gluten peptides, cross-linking gliadin with interstitial matrix proteins, and regulating RhoA activation. This last role is probably also performed by MYO9B, with as tTG has an influence on the intestinal barrier formed by the enterocytes in the intestine. This intestinal barrier is impaired in CD patients and loses some of its ability to regulate both the passage of gluten peptides and other molecules, and the protection against pathogens. Recent observations as reviewed in Jabri et al. have shown that not only the adaptive immune response is involved in CD, but also the innate immune response, and that the different gluten peptides can have their own effect on either of these immune responses (89).

Observations from functional studies have shown that the changes in the intestines of CD patients are not just due to damage but also to a deregulation of the proliferation/differentiation ratio of the enterocytes. We have learnt much about the pathological mechanisms in CD over the last 30 years, but still cannot answer all the questions. There are still several black boxes to be discovered, while the players known to be involved in CD might still be hiding some functions that influence the pathology. A combination of studies is needed in order to define how these players act together leading to CD. Genetic studies will reveal many of the small genetic players in CD in the coming years, given the increasing availability of high-throughput methods. A greater knowledge of pathway analyses and the function of variants in the DNA will also play an important role. Besides this, functional studies will be needed to reveal the role of these genetic variants

in CD and to show which genes are only involved in the pathology but not in the genetic susceptibility to CD. These genes may be a consequence, or an enhancer, of the disease process rather than a cause; for example, the expression studies have revealed there are many genes involved in the changes in structure of the intestine, but most of these will not be genetically associated to CD (60,99). We should not forget the overlap seen between several autoimmune disorders, both clinically and overlapping linkage regions. The genes found in autoimmune disorders may play a role in multiple disorders (for example, the *CTLA-4* and *PTPN22* genes, although they do not play as clear a role in CD as in other disorders) (21–27,93,94). The linkage region found in chromosome 6q21 (opposite arm of the HLA region) might also harbor an immune-related gene since this region has been identified in multiple immune disorders including CD. There are also several disorders where the permeability of the epithelial barrier is impaired and an overlap in disease genes is seen.

The major environmental factor provoking CD is gluten, but other, less important, environmental factors must also play a role, and the search for them has hardly begun. New ideas about pathology will also come from a clinical point of view; for example, clinicians showed years ago that there was an increased intestinal permeability. This is only now being linked to genes involved in CD. Clinicians should also help in defining the criteria for diagnosing a CD patient. For a long time, the diagnosis was made by a biopsy showing Marsh III pathology, but the question as to whether individuals with a Marsh I or Marsh II biopsy are, in fact, also CD patients is arising more and more. Perhaps genetic factors will be used for diagnoses, but so far the only genes playing a large role is are the *HLA-DQA1* and *-DQB1* and they are used to exclude the possibility of CD. Our main conclusion from the current genetic and functional studies is that we should look for causal genes in the barrier function as well as in the immune system.

Box 1: Challenges in finding genes in complex genetic disorders, with *MYO9B* as an example

Several factors influence our ability to find the genes involved in CD and to replicate them in multiple populations of CD patients. Since CD is not a monogenic disorder but a complex genetic disorder, multiple genes as well as environmental factors must play a role. The disease-causing variants will most often be common ones, and each on its own will not be sufficient to cause disease (100–103) because they will have a low RR. These variants have to occur frequently enough in the general population for them to co-occur in the individuals affected by these variants. Not every patient needs to have exactly the same combination of disease-predisposing variants, so this genetic heterogeneity adds to the difficulty of finding causative genes. In addition, healthy individuals can also harbor several disease-predisposing gene variants, but not enough or not in the right combination, to cause the disorder. The fact that genetic heterogeneity occurs, that the RR of the variants is low, and that the control population will also harbor these genetic variants, although in a smaller number, means it is necessary to have a large, well defined patient group, as well as a large control group to perform genetic studies.

There are several strategies for searching for disease susceptibility genes, like candidate gene/pathway studies and genome-wide studies. Both strategies can be studied in families using linkage-based designs, and in populations using case-control designs. Candidate gene/pathway studies require biological knowledge about the genes involved, but if they result in an association they often lead to a susceptibility gene being identified. Genome-wide studies, on the other hand, are hypothesis-free but when applied to a family-based linkage design the linkage regions mostly contain ~50–200 genes. Both strategies have their strengths and weaknesses, and the new possibility of whole-genome association studies will combine some of the strengths of both studies, since no biological knowledge is required and a positive result will lead to a single, associated gene. These studies will, of course, not avoid all the problems like population stratification, power issues and multiple testing, and a positive replication study will always be needed.

MYO9B as an example

The most recent success in CD research was the finding of *MYO9B* as a gene causing susceptibility to CD. A family study in the Dutch population found linkage to chromosome 19p13 (MYO9B, previously called the CELIAC4 locus) with an estimated λ_s of 2.6 (104). Subsequently, an association study was performed using microsatellite markers, which

resulted in association to a microsatellite marker in intron 1 of *MYO9B* with CD. Further single nucleotide polymorphism (SNP) typing in and around *MYO9B* showed association to the linkage disequilibrium (LD) block that covered the 3' site of the gene (40). A comprehensive tagging screen using SNPs in the whole MYO9B locus showed no association to other parts of this locus. *MYO9B* was found in the Dutch population with an estimated RR of 1.7 for heterozygous carriers of the associated variant and an RR of 2.3 for homozygous carriers. The expected population attributable risk of this variant is 23%–25%, meaning that removing this risk factor from the population reduces the risk to CD with 23%–25%. Several questions remain even though the gene responsible for the MYO9B locus has now been identified. First of all, what is the causative variant? The variants found associated to CD are located in a block with high LD and might just be tagging the real causative variant. Perhaps multiple rare variants together cause the association, as seen for the *CARD15* gene in IBD (105). Secondly, what is the exact RR and does this explain the observed linkage? Most genetic studies suffer from the 'winner's curse' and they tend to overestimate the RR (43). This could mean that MYO9B does not explain all the linkage and may imply that a second gene should be found even though the comprehensive study in the linkage region provided no evidence for this. If we take the lower boundary of the 95% confidence interval for the odds ratio, the RR would be 1.23, or perhaps even lower in reality. The chance that a replication study will have sufficient power to observe association to CD in a different population is very low. This could explain the negative results from the replication studies of Hunt et al. and Admundsen et al. (41,42). The first study had a cohort size that came close to the original Dutch cohort, but the second study was performed in a family setting, which is less prone to population stratification but also less powerful, and a case-control setting using smaller cohorts than the original study. Both were underpowered if we assume a low RR. The recent observation that MYO9B is also involved in IBD (especially in ulcerative colitis), with an OR of 1.2, could indeed indicate that the studies' power was too low (63). Another explanation could be that the original Dutch findings were false-positive, but we consider the chance of that low given that our result was found in two separate case-control groups. Since the causative variant has not been uncovered, it is possible that the tag SNPs used were not in complete LD with the causative variant, and there might have been slightly different LD patterns between these three studies. If this proves to be true, then testing the causative variant (once it has been found), in the other two populations might show that *MYO9B* is associated in multiple CD populations after all.

Acknowledgements

We thank Erica van Oort, Jackie Senior and Alexandra Zhernakova for critically reading the manuscript. The authors were supported by grants from the Netherlands Organization for Scientific Research (grant 912-02-028), and the Celiac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch Government (grant BSIK03009).

Conflict of interest

The authors declare no conflict of interest.

References

1. Rewers M. Epidemiology of celiac disease: what are the prevalence, incidence, and progression of celiac disease? Gastroenterology. 2005;128(Suppl 1):S47–51.
2. Dewar DH, Ciclitira PJ. Clinical features and diagnosis of celiac disease. Gastroenterology. 2005;128:S19–S24.
3. Rampertab SD, Pooran N, Brar P, Singh P, Green PH. Trends in the presentation of celiac disease. Am J Med. 2006;119:355.e9–14.
4. Wahab PJ, Meijer JW, Mulder CJ. Histologic follow-up of people with celiac disease on a gluten-free diet: slow and incomplete recovery. Am J Clin Pathol. 2002;118:459–63.
5. Schulzke JD, Bentzel CJ, Schulzke I, Riecken EO, Fromm M. Epithelial tight junction structure in the jejunum of children with acute and treated celiac sprue. Pediatr Res. 1998;43(Pt 1):435–41.
6. Schulzke JD, Schulzke I, Fromm M, Riecken EO. Epithelial barrier and ion transport in coeliac sprue: electrical measurements on intestinal aspiration biopsy specimens. Gut. 1995;37:777–82.
7. Uil JJ, van Elburg RM, van Overbeek FM, Meyer JW, Mulder CJ, Heymans HS. Follow-up of treated coeliac patients: sugar absorption test and intestinal biopsies compared. Eur J Gastroenterol Hepatol. 1996;8:219–23.
8. Smecuol E, Sugai E, Niveloni S, Vazquez H, Pedreira S, Mazure R, et al. Permeability, zonulin production, and enteropathy in dermatitis herpetiformis. Clin Gastroenterol Hepatol. 2005;3:335–41.
9. Zone JJ. Skin manifestations of celiac disease. Gastroenterology. 2005;128(Suppl 1):S87–91.
10. Daum S, Cellier C, Mulder CJJ. Refractory coeliac disease. Best Pract Res Clin Gastroenterol. 2005;19:413–24.

11. Green PH, Jabri B. Celiac disease. Annu Rev Med. 2006;57:207–21.
12. Schuppan D, Dieterich W, Ehnis T, Bauer M, Donner P, Volta U, et al. Identification of the Autoantigen of Celiac Disease. Ann NY Acad Sci. 1998;859:121–6.
13. 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. Marsh M. 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.
15. 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.
16. Dube C, Rostom A, Sy R, Cranney A, Saloojee N, Garritty C, et al. The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. Gastroenterology. 2005;128(Suppl 1):S57–67.
17. Fasano A, Berti I, Gerarduzzi T, Not T, Colletti RB, Drago S, 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.
18. Talley NJ, Valdovinos M, Petterson TM, Carpenter HA, Melton LJ 3rd. Epidemiology of celiac sprue: a community-based study. Am J Gastroenterol. 1994;89:843–6.
19. Greco L, Romino R, Coto I, Di Cosmo N, Percopo S, Maglio M, et al. The first large population based twin study of coeliac disease. Gut. 2002;50:624–8.
20. Nistico L, Fagnani C, Coto I, Percopo S, Cotichini R, Limongelli MG, et al. Concordance, disease progression, and heritability of coeliac disease in Italian twins. Gut. 2006;55:803–8.
21. King AL, Moodie SJ, Fraser JS, Curtis D, Reid E, Dearlove AM, 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.
22. Mora B, Bonamico M, Indovina P, Megiorni F, Ferri M, Carbone MC, et al. CTLA-4 +49 A/G dimorphism in Italian patients with celiac disease. Hum Immunol. 2003;64: 297–301.
23. van Belzen MJ, Mulder CJ, Zhernakova A, Pearson PL, Houwen RH, Wijmenga C. CTLA4 +49 A/G and CT60 polymorphisms in Dutch coeliac disease patients. Eur J Hum Genet. 2004;12:782–5.
24. Clot F, Fulchignoni-Lataud MC, Renoux C, Percopo S, Bouguerra F, Babron MC, 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.
25. 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.
26. Holopainen P, Naluai AT, Moodie S, Percopo S, Coto I, Clot F, et al. Candidate gene region 2q33 in European families with coeliac disease. Tissue Antigens. 2004;63: 212–22.
27. van Heel DA, Hunt K, Greco L, Wijmenga C. Genetics in coeliac disease. Best Pract Res Clin Gastroenterol. 2005;19:323–39.
28. Falchuk ZM, Rogentine GN, Strober W. Predominance of histocompatibility antigen HL-A8 in patients with gluten-sensitive enteropathy. J Clin Invest. 1972;51:1602–5.
29. Falchuk ZM, Strober W. HL-A antigens and adult coeliac disease. Lancet. 1972;2:1310.
30. Stokes PL, Asquith P, Holmes GK, Mackintosh P, Cooke WT. Histocompatibility antigens associated with adult coeliac disease. Lancet. 1972;2:162–4.
31. Corazza GR, Tabacchi P, Frisoni M, Prati C, Gasbarrini G. DR and non-DR Ia allotypes are associated with susceptibility to coeliac disease. Gut. 1985;26:1210–3.
32. Keuning JJ, Pena AS, van Leeuwen A, van Hooff JP, va Rood JJ. HLA-DW3 associated with coeliac disease. Lancet. 1976;1:506–8.
33. Tosi R, Vismara D, Tanigaki N, Ferrara GB, Cicimarra F, Buffolano W, et al. Evidence that celiac disease is primarily associated with a DC locus allelic specificity. Clin Immunol Immunopathol. 1983;28:395–404.
34. Kagnoff MF, Harwood JI, Bugawan TL, Erlich HA. Structural analysis of the HLA-DR, -DQ, and -DP alleles on the celiac disease-associated HLA-DR3 (DRw17) haplotype. Proc Natl Acad Sci U S A. 1989;86:6274–8.
35. Rittner C, DeMarchi M, Mollenhauer E, Carbonara A. Coeliac disease and C4A*QO: an association secondary to HLA-DR3. Tissue Antigens. 1984;23:130–4.
36. Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E. Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. J Exp Med. 1989;169:345–50.
37. Tosi R, Tanigaki N, Polanco I, De Marchi M, Woodrow JC, Hetzel PA. A radioimmunoassay typing study of non-DQw2-associated celiac disease. Clin Immunol Immunopathol. 1986;39:168–72.
38. Bevan S, Popat S, Braegger CP, Busch A, O'Donoghue D, Falth-Magnusson K, et al. Contribution of the MHC region to the familial risk of coeliac disease. J Med Genet. 1999;36:687–90.
39. Petronzelli F, Bonamico M, Ferrante P, Grillo R, Mora B, Mariani P, et al. Genetic contribution of the HLA region to the familial clustering of coeliac disease. Ann Hum Genet. 1997;61(Pt 4):307–17.
40. Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, et al. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. Nat Genet. 2005;37: 1341–4.
41. Amundsen SS, Monsuur AJ, Wapenaar MC, Lie BA, Ek J, Gudjonsdottir AH, et al. Association Analysis of MYO9B Gene Polymorphisms with Celiac Disease in a Swedish/ Norwegian Cohort. Hum Immunol. 2006;67:341–5.
42. Hunt KA, Monsuur AJ, McArdle W, Kumar PJ, Travis SP, Walters JR, et al. Lack of association of MYO9B genetic variants with coeliac disease in a British cohort. Gut. 2006;55:969–72.
43. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet. 2001;29:306–9.
44. van Herpen TW, Goryunova SV, van der Schoot J, Mitreva M, Salentijn E, Vorst O, et al. Alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes. BMC Genomics. 2006;7:1.
45. Koning F, Gilissen L, Wijmenga C. Gluten: a two-edged sword. Immunopathogenesis of celiac disease. Springer Semin Immunopathol. 2005;27:217–32.
46. Molberg O, Uhlen AK, Jensen T, Flaete NS, Fleckenstein B, Arentz-Hansen H, et al. Mapping of gluten T-cell epitopes in the bread wheat ancestors: implications for celiac disease. Gastroenterology. 2005;128:393–401.

47. Spaenij-Dekking L, Kooy-Winkelaar Y, van Veelen P, Drijfhout JW, Jonker H, van Soest L, et al. Natural variation in toxicity of wheat: potential for selection of nontoxic varieties for celiac disease patients. Gastroenterology. 2005;129:797–806.
48. Spaenij-Dekking L, Kooy-Winkelaar Y, Koning F. The Ethiopian cereal tef in celiac disease. N Engl J Med. 2005;353:1748–9.
49. Hausch F, Shan L, Santiago NA, Gray GM, Khosla C. Intestinal digestive resistance of immunodominant gliadin peptides. Am J Physiol Gastrointest Liver Physiol. 2002;283:G996–G1003.
50. Vanhoof G, Goossens F, Hendriks L, De Meester I, Hendriks D, Vriend G, et al. Cloning and sequence analysis of the gene encoding human lymphocyte prolyl endopeptidase. Gene. 1994;149:363–6.
51. Higaki-Sato N, Sato K, Esumi Y, Okumura T, Yoshikawa H, Tanaka-Kuwajima C, et al. Isolation and identification of indigestible pyroglutamyl peptides in an enzymatic hydrolysate of wheat gluten prepared on an industrial scale. J Agric Food Chem. 2003;51:8–13.
52. Shan L, Molberg O, Parrot I, Hausch F, Filiz F, Gray GM, et al. Structural basis for gluten intolerance in celiac sprue. Science. 2002;297:2275–9.
53. Matysiak-Budnik T, Candalh C, Dugave C, Namane A, Cellier C, Cerf-Bensussan N, et al. Alterations of the intestinal transport and processing of gliadin peptides in celiac disease. Gastroenterology. 2003;125:696–707.
54. Diosdado B, Stepniak DT, Monsuur AJ, Franke L, Wapenaar MC, Mearin ML, et al. No genetic association of the human prolyl endopeptidase gene in the Dutch celiac disease population. Am J Physiol Gastrointest Liver Physiol. 2005;289:G495–500.
55. Monsuur AJ, Stepniak D, Diosdado B, Wapenaar MC, Mearin ML, Koning F, et al. Genetic and functional analysis of pyroglutamyl-peptidase I in coeliac disease. Eur J Gastroenterol Hepatol. 2006;18:637–44.
56. Stepniak D, Spaenij-Dekking L, Mitea C, Moester M, de Ru A, Baak-Pablo R, et al. Highly efficient gluten degradation with a newly identified prolyl endoprotease: implications for celiac disease. Am J Physiol Gastrointest Liver Physiol. 2006;291:G621–9.
57. Watson AJ, Chu S, Sieck L, Gerasimenko O, Bullen T, Campbell F, et al. Epithelial barrier function in vivo is sustained despite gaps in epithelial layers. Gastroenterology. 2005;129:902–12.
58. Ismail AS, Hooper LV. Epithelial cells and their neighbors. IV. Bacterial contributions to intestinal epithelial barrier integrity. Am J Physiol Gastrointest Liver Physiol. 2005;289: G779–84.
59. Forsberg G, Fahlgren A, Horstedt P, Hammarstrom S, Hernell O, Hammarstrom ML. Presence of bacteria and innate immunity of intestinal epithelium in childhood celiac disease. Am J Gastroenterol. 2004;99:894–904.
60. Diosdado B, Wapenaar MC, Franke L, Duran KJ, Goerres MJ, Hadithi M, et al. A microarray screen for novel candidate genes in coeliac disease pathogenesis. Gut. 2004;53:944–51.
61. Mallant-Hent R, Mary B, von Blomberg E, Yuksel Z, Wahab PJ, Gundy C, et al. Disappearance of anti-Saccharomyces cerevisiae antibodies in coeliac disease during a gluten-free diet. Eur J Gastroenterol Hepatol. 2006;18:75–8.
62. Zolotarevsky Y, Hecht G, Koutsouris A, Gonzalez DE, Quan C, Tom J, et al. A membrane-permeant peptide that inhibits MLC kinase restores barrier function in in vitro models of intestinal disease. Gastroenterology. 2002;123: 163–72.
63. van Bodegraven AA, Curley CR, Hunt KA, Monsuur AJ, Linskens RK, Onnie CM, et al., Genetic variation in myosin IXB is associated with ulcerative colitis. Gastroenterology. Published online, Sept 1, 2006.
64. 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.
65. Post PL, Tyska MJ, O'Connell CB, Johung K, Hayward A, Mooseker MS. Myosin-IXb is a single-headed and processive motor. J Biol Chem. 2002;277:11679–83.
66. Matter K, Balda MS. Signalling to and from tight junctions. Nat Rev Mol Cell Biol. 2003;4:225–36.
67. Janiak A, Zemskov EA, Belkin AM. Cell surface transglutaminase promotes RhoA activation via integrin clustering and suppression of the Src-p190RhoGAP signaling pathway. Mol Biol Cell. 2006;17:1606–19.
68. Blair SA, Kane SV, Clayburgh DR, Turner JR. Epithelial myosin light chain kinase expression and activity are upregulated in inflammatory bowel disease. Lab Invest. 2006;86:191–201.
69. Russo JM, Florian P, Shen L, Graham WV, Tretiakova MS, Gitter AH, et al. Distinct temporal-spatial roles for rho kinase and myosin light chain kinase in epithelial purse-string wound closure. Gastroenterology. 2005;128: 987–1001.
70. Liu Z, Li N, Neu J. Tight junctions, leaky intestines, and pediatric diseases. Acta Paediatr. 2005;94:386–93.
71. Schreiber S. Slipping the barrier: how variants in CARD15 could alter permeability of the intestinal wall and population health. Gut. 2006;55:308–9.
72. Schreiber S, Rosenstiel P, Albrecht M, Hampe J, Krawczak M. Genetics of Crohn disease, an archetypal inflammatory barrier disease. Nat Rev Genet. 2005;6: 376–88.
73. Cookson W. The immunogenetics of asthma and eczema: a new focus on the epithelium. Nat Rev Immunol. 2004;4: 978–88.
74. Secondulfo M, Iafusco D, Carratu R, deMagistris L, Sapone A, Generoso M, et al. Ultrastructural mucosal alterations and increased intestinal permeability in non-celiac, type I diabetic patients. Dig Liver Dis. 2004;36: 35–45.
75. Karell K, Louka AS, Moodie SJ, Ascher H, Clot F, Greco L, 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.
76. Margaritte-Jeannin P, Babron MC, Bourgey M, Louka AS, Clot F, Percopo S, et al. HLA-DQ relative risks for coeliac disease in European populations: a study of the European Genetics Cluster on Coeliac Disease. Tissue Antigens. 2004;63:562–7.
77. Vader W, Stepniak D, Kooy Y, Mearin L, Thompson A, van Rood JJ, et al. The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. Proc Natl Acad Sci U S A. 2003;100:12390–5.
78. Koning F. Celiac disease: caught between a rock and a hard place. Gastroenterology. 2005;129:1294–301.
79. Al-Toma A, Goerres MS, Meijer JW, Pena AS, Crusius JB, Mulder CJ. Human leukocyte antigen-DQ2 homozygosity and the development of refractory celiac disease and enteropathy-associated T-cell lymphoma. Clin Gastroenterol Hepatol. 2006;4:315–9.

80. Kim CY, Quarsten H, Bergseng E, Khosla C, Sollid LM. Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. Proc Natl Acad Sci U S A. 2004;101:4175–9.
81. Molberg O, McAdam SN, Korner R, Quarsten H, Kristiansen C, Madsen L, 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.
82. Qiao SW, Bergseng E, Molberg O, Jung G, Fleckenstein B, Sollid LM. Refining the rules of gliadin T cell epitope binding to the disease-associated DQ2 molecule in celiac disease: importance of proline spacing and glutamine deamidation. J Immunol. 2005;175:254–61.
83. Vader LW, de Ru A, van der Wal Y, Kooy YMC, Benckhuijsen W, Mearin ML, et al. Specificity of Tissue Transglutaminase Explains Cereal Toxicity in Celiac Disease. J Exp Med. 2002;195:643–9.
84. Ciccocioppo R, Di Sabatino A, Ara C, Biagi F, Perilli M, Amicosante G, et al. Gliadin and tissue transglutaminase complexes in normal and coeliac duodenal mucosa. Clin Exp Immunol. 2003;134:516–24.
85. Sakly W, Sriha B, Ghedira I, Bienvenu F, Ayadi A, Sfar MT, et al. Localization of tissue transglutaminase and N (epsilon)-(gamma) -glutamyl lysine in duodenal cucosa during the development of mucosal atrophy in coeliac disease. Virchows Arch. 2005;446:613–8.
86. Dieterich W, Esslinger B, Trapp D, Hahn E, Huff T, Seilmeier W, et al. Cross linking to tissue transglutaminase and collagen favours gliadin toxicity in coeliac disease. Gut. 2006;55:478–84.
87. 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.
88. 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.
89. Jabri B, Kasarda DD, Green PH. Innate and adaptive immunity: the yin and yang of celiac disease. Immunol Rev. 2005;206:219–31.
90. Anderson RP, Degano P, Godkin AJ, Jewell DP, Hill AV. In vivo antigen challenge in celiac disease identifies a single transglutaminase-modified peptide as the dominant A-gliadin T-cell epitope. Nat Med. 2000;6:337–42.
91. Arentz-Hansen H, Korner R, Molberg O, Quarsten H, Vader W, Kooy YM, et al. The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. J Exp Med. 2000;191:603–12.
92. Diosdado B, Wijmenga C. Molecular mechanisms of the adaptive, innate and regulatory immune responses in the intestinal mucosa of celiac disease patients. Expert Rev Mol Diagn. 2005;5:681–700.
93. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet. 2004;36:337–8.
94. Zhernakova A, Eerligh P, Wijmenga C, Barrera P, Roep BO, Koeleman BP. Differential association of the PTPN22 coding variant with autoimmune diseases in a Dutch population. Genes Immun. 2005;6:459–61.
95. Hue S, Mention JJ, Monteiro RC, Zhang S, Cellier C, Schmitz J, et al. A direct role for NKG2D/MICA interaction in villous atrophy during celiac disease. Immunity. 2004;21:367–77.
96. van Belzen MJ, Koeleman BP, Crusius JB, Meijer JW, Bardoel AF, Pearson PL, et al. Defining the contribution of the HLA region to cis DQ2-positive coeliac disease patients. Genes Immun. 2004;5:215–20.
97. Di Sabatino A, Ciccocioppo R, Cupelli F, Cinque B, Millimaggi D, Clarkson MM, et al. Epithelium derived interleukin 15 regulates intraepithelial lymphocyte Th1 cytokine production, cytotoxicity, and survival in coeliac disease. Gut. 2006;55:469–77.
98. Meresse B, Chen Z, Ciszewski C, Tretiakova M, Bhagat G, Krausz TN, et al. Coordinated induction by IL15 of a TCR-independent NKG2D signaling pathway converts CTL into lymphokine-activated killer cells in celiac disease. Immunity. 2004;21:357–66.
99. Juuti-Uusitalo K, Maki M, Kaukinen K, Collin P, Visakorpi T, Vihinen M, et al. cDNA microarray analysis of gene expression in coeliac disease jejunal biopsy samples. J Autoimmun. 2004;22:249–65.
100. Chakravarti A. Population genetics—making sense out of sequence. Nat Genet. 1999;21(Suppl):56–60.
101. Glazier AM, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. Science. 2002;298:2345–9.
102. Risch NJ. Searching for genetic determinants in the new millennium. Nature. 2000;405:847–56.
103. Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. Nat Rev Genet. 2004;5:89–100.
104. Van Belzen MJ, Meijer JW, Sandkuijl LA, Bardoel AF, Mulder CJ, Pearson PL, et al. A major non-HLA locus in celiac disease maps to chromosome 19. Gastroenterology. 2003;125:1032–41.
105. Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, et al. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet. 2001;29:223–8.

Chapter 1.2

MYO9B and celiac disease

Alienke J. Monsuur & Cisca Wijmenga

Myosin IXB

General introduction

Recently the *myosin IXB* (*MYO9B*) gene on chromosome 19 was shown to be involved in celiac disease (1). Not a lot is known about the function of MYO9B in general and the role and importance of MYO9B in celiac disease needs to be elucidated.

In this chapter we will discuss the literature on MYO9B to examine its function, and combine this with the knowledge of celiac disease to form hypotheses about the functional role of MYO9B in celiac disease.

Myosins

MYO9B belongs to the myosin superfamily that comprises 40 genes that can roughly be divided into two groups, the conventional myosins and the unconventional myosins (2, 3). The conventional myosins are mostly muscle myosins whereas the unconventional myosins (among which is MYO9B) are seen more in non-muscle cells, where up to ten different myosins can be expressed in one cell. The myosins are divided into 11 classes in humans and each myosin has a heavy chain with a conserved ~80 kDA catalytic domain. Some of the myosins also have an α-helical light chain-binding region consisting of one or more IQ motifs and a C-terminal tail and/or an N-terminal extension. It is this tail domain that leads to class-specific functions. Many myosins bind to or localize at the cell membranes. They are involved in all kind of actin dynamics – e.g. spatial and temporal organization, like actin filament nucleation and elongation, movement and transport of organelles along actin filaments, and signal transduction (4-6). Their importance is shown by the fact that genetic mutations in some unconventional myosins lead to deafness (class VI, VII and possibly class I), blindness (class III and VII) and seizures (class V) (reviewed in Mermal) (6).

The *MYO9B* gene and expression pattern

The *MYO9B* gene consists of 40 exons, spanning 7510 bp and is translated into a protein of 2157 amino acids according to the Ensembl database (release 42, December 2006) (see Figure 1). No alternative transcripts are known in the Ensembl database. Our group looked at the expression levels of MYO9B in human immature and mature dendritic cells, human CD4-positive T-cells of two different donors, the HL60 leukocyte cell line and Jurkat T-cell line (see Figure 2). A study by Wirth et al. showed high expression of MYO9B in peripheral blood leukocytes, moderate expression in spleen and thymus, and minor levels in testis, prostate, ovary, brain, small intestine and lung tissue (7).

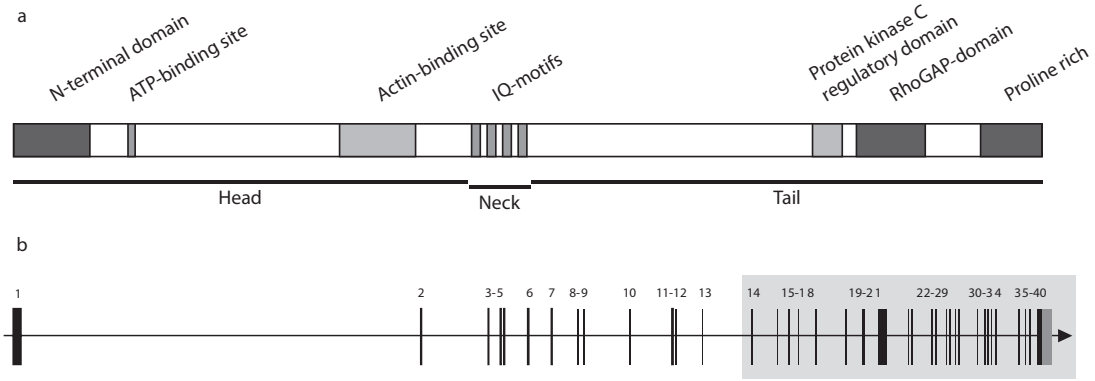


Figure 1 The myosin IXB gene. a) The domains and the head, neck and tail (after Wirth et al.). b) Exons and the associated region (gray area). IQ-motifs are located in exon 20 and 21, RhoGAP domain is located in exon 31-35.

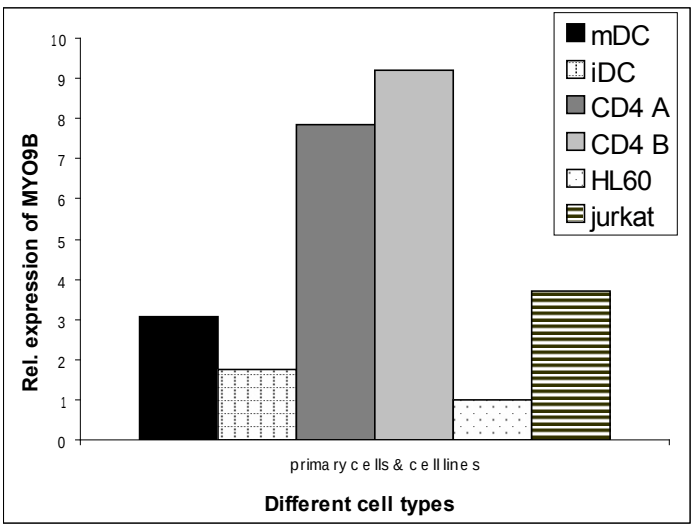


Figure 2. Expression data of MYO9B in primary human cells and HL60 and Jurkat T-cell lines. Expression of HL60 was set to 1, all other values were relative to the expression of HL60

MYO9B and its domains

MYR5 was found in rat, and homology searches led to the discovery of MYO9B in man and also of myo9b in mouse (7-11). MYO9B belongs to the class IX myosins, together with MYO9a (12). All myosins, including MYO9B, can be divided into three parts: a head, neck and tail. MYO9B has a head, which is the motor, of ~230 kDa with two insertions in the DNA sequence of the head that are not observed in other myosins, a 60-amino acid N-terminal extension and a 120-amino acid insert, the latter is within the contact site of MYO9B with actin (Figure 1) (13). O'Connell showed

that MYO9B moves towards the plus end of the actin filaments (14). Inoue et al. used a protein lacking a domain in the tail which moved in the opposite direction, leading to the suggestion that the tail determines the direction in which MYO9B moves (14, 15). The neck contains four IQ regions (named for the presence of tandem isoleucine and glutamine residues), which should be able to bind four light chains of the calmodulin/EF-hand superfamily (7, 8). These bound chains and the IQ regions form a mechanical lever used to move the myosin along actin filaments, the more IQ regions the longer the lever arm and step-size (16). The bound calmodulin light chains also influence the chemo-mechanical properties, leading to regulation of the motor for example (17, 18). As seen in some other myosin classes, the tail of MYO9B lacks the coiled coil region and MYO9B therefore functions as a single headed motor, unlike several other myosins that form double headed motors (19). Tail domains as seen in MYO9B are generally seen in proteins involved in signal transduction. It its tail, MYO9B has a protein kinase C regulatory domain. Besides this domain, the main feature of MYO9B that distinguishes the class IX myosins from the other classes is the RhoGAP-domain in its tail (GTPase activating protein (GAP) domain) (Figures 1 and 3a). This GTPase domain works on the members of the rho family of small GTP-binding proteins by inhibiting the GAP activity of Rho (Figure 3b) (18). The rho family GTPases are one of the factors involved in actin dynamics (4). A gene found to interact with MYO9B is Brefeldin A-inhibited guanine nucleotide exchange protein 1 (BIG1) (20), which belongs to the guanine nucleotide exchange proteins (GEFs). BIG1 promotes the exchange from GDP to GTP while the RhoGAP domain promotes the exchange from GTP to GDP. BIG1 competes with the GTPase activating protein RhoA for binding to MYO9B and could regulate the influence of MYO9B on RhoA. BIG1 is localized at the golgi and this might influence the regulation of MYO9B spatially (Figure 3c). Overall we can assume that MYO9B is carrying its own cargo, the Rho-GAP domain, to the site of action using the motor and the acting filaments as its transporter.

The different roles of MYO9B in celiac disease pathogenesis

MYO9B is a protein that carries its own cargo to the site of action. The step size is controlled via the IQ motifs, which also regulates the motor activity, and the tail influences the direction in which MYO9B moves. The RhoGAP domain can negatively regulate the activity of RhoA while the competition between RhoA and BIG1 can influence this regulation and the spatial organization of MYO9B in the cell. Based on these characteristics of MYO9B we have formed several hypotheses on its role in celiac disease pathogenesis (see below and Table 1).

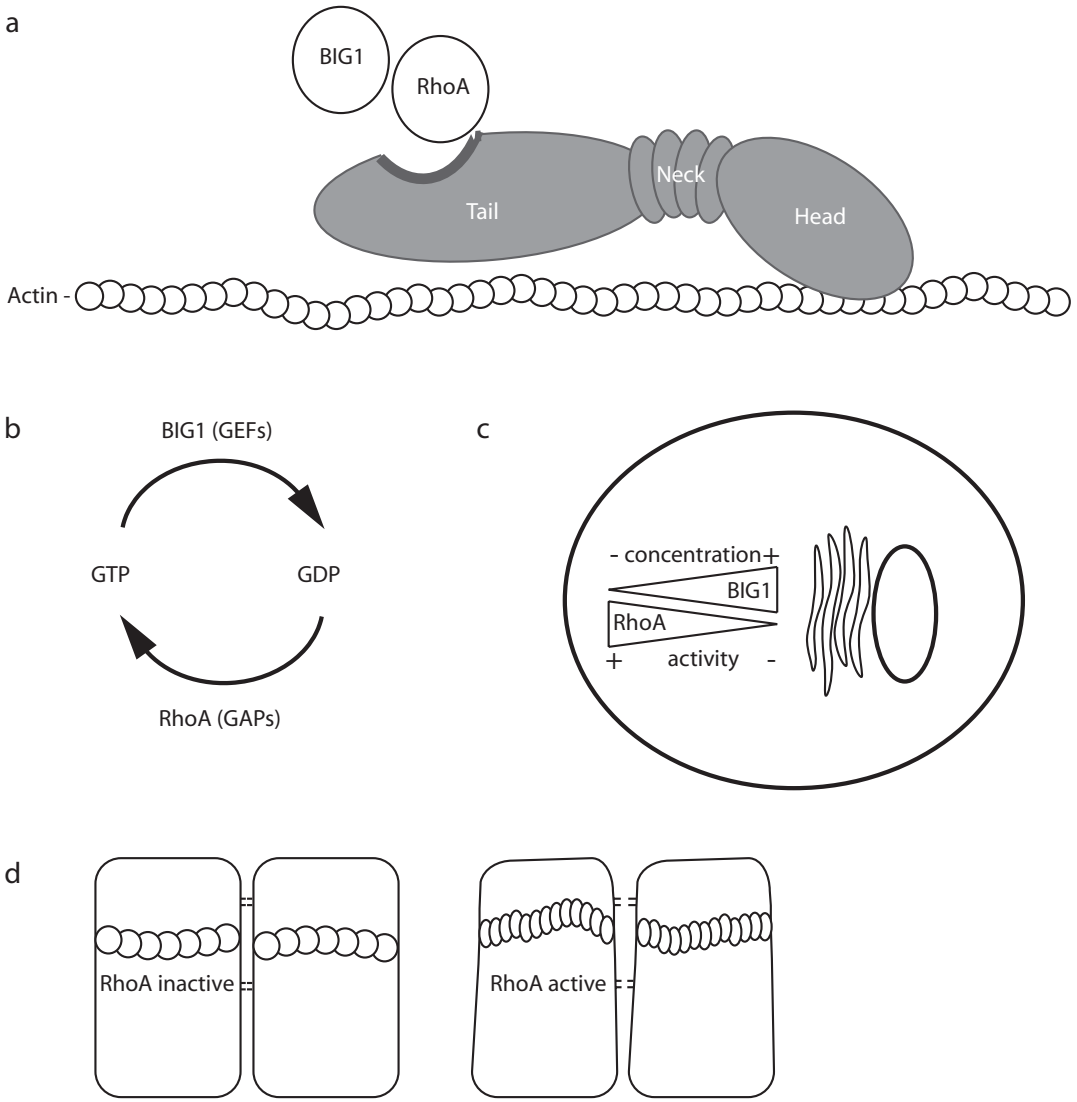


Figure 3. Schematic representations of the MYO9B protein and its action. a) MYO9B contains a head, neck and tail domain. It moves along actin filaments and can bind with its RhoGAP domain to the rho family of GTPases. BIG1 also binds MYO9B and competes with rho for the binding. b) BIG1 and rho have opposite actions. c) The high concentration of BIG1 around the Golgi might influence the activity of rho due to the competition in binding to MYO9B. d) Active rho leads to increased contraction of actin, and more permeable tight junctions

MYO9B is able to affect the rho family of GTPases and therefore the control of tight junction functioning and cytoskeletal modifications, leading to an increased permeability of the intestinal barrier.

Small GTPases like RhoA influence the tight junction assembly. The activity of rho should be balanced so that the barrier is not too tight (leukocytes and some molecules need to be able to pass through the barrier) and not too weak (the barrier needs to protect the body against pathogens and bacteria) (21-23). Functional changes in tight junction have been seen for Crohn's disease and studies in celiac disease showed different expression, localization and phosphorylation of tight junction proteins (24, 25). Via rho, MYO9B could alter the actin contraction leading to less efficient tight junctions and increased permeability of the barrier, with a less efficient defense by the barrier (Figure 3d).

Tight junctions not only regulate epithelial permeability but are also involved in the switch from proliferation to differentiation, since contact inhibition is needed for the epithelial cells in order to start differentiating (21). So MYO9B could, via rho, the adherens junction and the tight junctions, also alter the balance of proliferating and differentiating cells needed to maintain a good barrier.

MYO9B's motor activity is influenced via the IQ motifs resulting in a different velocity of MYO9B.

The key component in the MYO9B gene is the fact that it carries its own cargo to the site of action. The IQ motifs are involved in the step size and length of MYO9B, so alterations in the IQ motifs, due to genetic mutations, could alter this (16, 17) and MYO9B would not be able to arrive at the right place at the right time, so that its influence on rho would be less effective.

Different function of MYO9B influences the cytoskeletal rearrangements of the cells and their potential for bacteria invasion

Shigella infection of the intestine induces cytoskeletal rearrangements with a resulting invasion of bacterias into the epithelial cells. Graf et al. studied the role of myr5 (the rat form of MYO9B) in Shigella infection since rho is involved in this cytoskeletal rearrangement and MYO9B/myr5 influences the activity of rho (26). Overexpression of myr5 decreases the activity of rho and therefore diminishes the possibility of Shigella to invade the cells. This altered invasiveness of enterocytes can lead to differences in uptake of gluten into the enterocytes. It is not known how gluten enters cells or passes between cells, but if gluten passes the epithelial layer via a route that is not designed to handle the gluten peptides, abnormal reactions towards gluten might well occur.

Table 1. Hypotheses on the role of MYO9B in celiac disease

Role of MYO9B	Cell type	Effect
Altered influence via rho on tight junction function and cytoskeletal modification	Enterocyte	-Increased permeability of the intestinal barrier -Altered balance in switch from proliferating to differentiating cells
Different motor activity resulting in altered velocity of MYO9B influencing MYO9Bs spatial and temporal action	Enterocyte or lymphocyte	-Increased permeability of the intestinal barrier -Altered response of lymphocytes towards gluten
Influence on the cytoskeletal rearrangements leading to altered invasiveness of the cells	Enterocyte	-Different uptake of gluten into the enterocytes, via a route that might not be well equipped to handle gluten
Altered purse-string wound closure	Enterocyte or lymphocyte	-Increased vulnerability of the intestinal barrier -Altered response of lymphocytes towards gluten, or altered movement
Alterations in plasma membrane morphology leading to altered endocytosis	Enterocyte	-Different uptake of gluten into the enterocytes, via a route that might not be well equipped to handle gluten

Via rho, MYO9B has an influence on the MLCK/Rock purse-string wound closure of human epithelial intestinal wounds

The intestinal barrier is constantly being affected by local damage, like trauma, inflammation and shedding of apoptotic cells. The small wounds that occur are closed via a process of acto-myosin ring assembly and contraction of this ring (27). The ring assembly is triggered by Rho kinase (ROCK) a downstream effector of rho, the subsequent contraction of this ring is under the control of myosin light chain (MLC) and its kinase (MLCK). An altered purse-string wound closure leads to a more vulnerable epithelial layer.

The downstream effectors of ROCK influence the cytoskeleton of lymphocytes. Different cytoskeletal behavior of these cells can alter the T cell synapse formation, the migration of neutrophils, the diapedesis of monocytes out of the blood vessels or the ability of dendritic cells to sample molecules in the intestine through the epithelial layer. This would influence responses to gluten or bacteria.

MYO9B has an influence on endocytosis, leading to differences in how substances enter via the cell into the body.

It has been shown that MYR5 is partially membrane associated and that it goes to the bacteria entry spots during Shigella infection (discussed above) (26). MYOM from *D. discoideum*, a protein with an antagonistic effect to MYO9B has been seen on endosomal organelles and stimulation leads to alterations in the plasma membrane morphology (4). Besides this, class I myosins are involved in actin dynamics and endocytosis. Although MYO9B is not a class I myosin, it is involved in actin dynamics and it may well be involved in endocytosis. This would not be surprising since rhoA has a role in endocytosis and the lysosomal cycle (28). If this proves to be the case, then an impaired MYO9B could influence the endocytotic uptake of proteins, e.g. of gluten, leading to an altered entering of gluten into the body, which might lead to problems in the proper handling of gluten.

Overall most of the hypotheses on the role of MYO9B lead to effects on the barrier function of the enterocytes. Gluten is then able to enter the body in altered ways which might be less efficient in giving a proper response to or in breaking down the gluten. The second possible role of MYO9B is via an altered response of lymphocytes towards gluten or bacteria.

References

1. Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, et al. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. Nat Genet. 2005 Dec;37(12):1341-4.
2. Berg JS, Powell BC, Cheney RE. A millennial myosin census. Mol Biol Cell. 2001 Apr;12(4):780-94.
3. Oliver TN, Berg JS, Cheney RE. Tails of unconventional myosins. Cell Mol Life Sci. 1999 Oct 15;56(3-4):243-57.
4. Soldati T. Unconventional myosins, actin dynamics and endocytosis: a menage a trois? Traffic. 2003 Jun;4(6):358-66.
5. Bridgman PC, Elkin LL. Axonal myosins. J Neurocytol. 2000 Nov-Dec;29(11-12):831-41.
6. Mermall V, Post PL, Mooseker MS. Unconventional myosins in cell movement, membrane traffic, and signal transduction. Science. 1998 Jan 23;279(5350):527-33.
7. 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 Mar;109 (Pt 3):653-61.
8. Reinhard J, Scheel AA, Diekmann D, Hall A, Ruppert C, Bahler M. A novel type of myosin implicated in signalling by rho family GTPases. Embo J. 1995 Feb 15;14(4):697-704.
9. Grewal PK, Jones AM, Maconochie M, Lemmers RJ, Frants RR, Hewitt JE. Cloning of the murine unconventional myosin gene Myo9b and identification of alternative splicing. Gene.

1999 Nov 29;240(2):389-98.
10. Whitmer JD, Koslovsky JS, Bahler M, Mercer JA. Chromosomal location of three unconventional myosin heavy chain genes in the mouse. Genomics. 1996 Dec 1;38(2):235-7.
11. Bahler M, Kehrer I, Gordon L, Stoffler HE, Olsen AS. Physical mapping of human myosin-IXB (MYO9B), the human orthologue of the rat myosin myr 5, to chromosome 19p13.1. Genomics. 1997 Jul 1;43(1):107-9.
12. Gorman SW, Haider NB, Grieshammer U, Swiderski RE, Kim E, Welch JW, et al. The cloning and developmental expression of unconventional myosin IXA (MYO9A) a gene in the Bardet-Biedl syndrome (BBS4) region at chromosome 15q22-q23. Genomics. 1999 Jul 15;59(2):150-60.
13. Bahler M. Are class III and class IX myosins motorized signalling molecules? Biochim Biophys Acta. 2000 Mar 17;1496(1):52-9.
14. O'Connell CB, Mooseker MS. Native Myosin-IXb is a plus-, not a minus-end-directed motor. Nat Cell Biol. 2003 Feb;5(2):171-2.
15. Inoue A, Saito J, Ikebe R, Ikebe M. Myosin IXb is a single-headed minus-end-directed processive motor. Nat Cell Biol. 2002 Apr;4(4):302-6.
16. Bahler M, Rhoads A. Calmodulin signaling via the IQ motif. FEBS Lett. 2002 Feb 20;513(1):107-13.
17. Post PL, Tyska MJ, O'Connell CB, Johung K, Hayward A, Mooseker MS. Myosin-IXb is a single-headed and processive motor. J Biol Chem. 2002 Apr 5;277(14):11679-83.
18. Post PL, Bokoch GM, Mooseker MS. Human myosin-IXb is a mechanochemically active motor and a GAP for rho. J Cell Sci. 1998 Apr;111 (Pt 7):941-50.
19. Kambara T, Ikebe M. A unique ATP hydrolysis mechanism of single-headed processive myosin, myosin IX. J Biol Chem. 2006 Feb 24;281(8):4949-57.
20. Saeki N, Tokuo H, Ikebe M. BIG1 is a binding partner of myosin IXb and regulates its Rho-GTPase activating protein activity. J Biol Chem. 2005 Mar 18;280(11):10128-34.
21. Matter K, Balda MS. Signalling to and from tight junctions. Nat Rev Mol Cell Biol. 2003 Mar;4(3):225-36.
22. Bruewer M, Hopkins AM, Hobert ME, Nusrat A, Madara JL. RhoA, Rac1, and Cdc42 exert distinct effects on epithelial barrier via selective structural and biochemical modulation of junctional proteins and F-actin. Am J Physiol Cell Physiol. 2004 Aug;287(2):C327-35.
23. Matter K, Aijaz S, Tsapara A, Balda MS. Mammalian tight junctions in the regulation of epithelial differentiation and proliferation. Curr Opin Cell Biol. 2005 Oct;17(5):453-8.
24. Ciccocioppo R, Finamore A, Ara C, Di Sabatino A, Mengheri E, Corazza GR. Altered expression, localization, and phosphorylation of epithelial junctional proteins in celiac disease. Am J Clin Pathol. 2006 Apr;125(4):502-11.
25. Soderholm JD, Olaison G, Peterson KH, Franzen LE, Lindmark T, Wiren M, et al. Augmented increase in tight junction permeability by luminal stimuli in the non-inflamed ileum of Crohn's disease. Gut. 2002 Mar;50(3):307-13.
26. Graf B, Bahler M, Hilpela P, Bowe C, Adam T. Functional role for the class IX myosin myr5 in epithelial cell infection by Shigella flexneri. Cell Microbiol. 2000 Dec;2(6):601-16.
27. Russo JM, Florian P, Shen L, Graham WV, Tretiakova MS, Gitter AH, et al. Distinct temporal-spatial roles for rho kinase and myosin light chain kinase in epithelial purse-string wound closure. Gastroenterology. 2005 Apr;128(4):987-1001.
28. Ellis S, Mellor H. Regulation of endocytic traffic by rho family GTPases. Trends Cell Biol. 2000 Mar;10(3):85-8.

Part 2

Enzymes involved in gluten breakdown

Chapter 2.1

Genetic and functional analysis of pyroglutamyl-peptidase I in coeliac disease

Alienke J. Monsuur, Dariusz Stepniak, Begoña Diosdado, Martin C. Wapenaar, Maria Luisa Mearin, Frits Koning and Cisca Wijmenga

European Journal of Gastroenterology & Hepatology 2006; 18:637–644

Genetic and functional analysis of pyroglutamyl-peptidase I in coeliac disease

Alienke J. Monsuur[a], Dariusz Stepniak[b], Begoña Diosdado[a,d], Martin C. Wapenaar[a], Maria Luisa Mearin[c], Frits Koning[b] and Cisca Wijmenga[a]

Coeliac disease (CD) is an enteropathy caused by an immune reaction towards wheat gluten and similar proteins from barley and rye. It was shown that some gluten peptides spontaneously form *N*-terminal L-pyroglutamate. This modification could potentially make gluten more resistant to proteolytic degradation within the intestine. Pyroglutamyl-peptidase I (PGPEPI) is an enzyme that hydrolytically removes the L-pyroglutamyl residues that render the modified proteins and peptides more sensitive to degradation by other proteases. Interestingly, we found that the *PGPEP1* gene is located in a CD susceptibility locus. As an impaired enzyme function caused by genetic alterations might increase the amount of immunogenic gluten peptides, we conducted a comprehensive functional genomics analysis of *PGPEP1*, including DNA sequencing, genetic association testing, and quantifying RNA expression. We also determined the enzymatic activity of PGPEPI in duodenal biopsies. Our results uniformly indicate that *PGPEP1* is not involved in the aetiology and pathology of CD. *Eur J Gastroenterol Hepatol* 18:637–644 © 2006 Lippincott Williams & Wilkins.

European Journal of Gastroenterology & Hepatology 2006, **18**:637–644

Keywords: coeliac disease, pyroglutamyl-peptidase I, candidate gene, genetic association, functional study

[a]Complex Genetics Section, DBG-Department of Medical Genetics, University Medical Centre, Utrecht, the Netherlands, [b]Departments of Immunohaematology and Blood Transfusion, [c]Paediatrics, Paediatric Gastroenterology Unit, Leiden University Medical Centre, Leiden, the Netherlands and [d]Department of Pathology, VU University Medical Centre, Amsterdam, the Netherlands

Correspondence to Cisca Wijmenga, Complex Genetics Section, Stratenum 2.117, Department of Biomedical Genetics, University Medical Centre Utrecht, PO Box 85060, 3508 AB Utrecht, the Netherlands
Tel: +31 30 253 8427; fax: +31 30 253 8479;
e-mail: t.n.wijmenga@med.uu.nl

Sponsorship: This study was supported by the Netherlands Organization for Health Research and Development (*ZonMw*, grants 902-22-094 and 912-02-028), the Coeliac Disease Consortium, an innovative cluster approved by the Netherlands Genomics Initiative and partly funded by a Dutch government grant (*BSIK*, 03009), and the Dutch Digestive Diseases Foundation (*MLDS*, grants WS00-13 and WS03-06) awarded to C.W. and F.K.

Received 17 November 2005 **Accepted** 6 March 2006

Introduction

Pyroglutamyl-peptidase I (PGPEPI; EC 3.4.19.3) is a cytosolic cysteine peptidase, which hydrolytically removes the terminal *N*-pyroglutamyl (L-pGlu) residue from proteins and peptides [1,2]. These L-pGlu residues confer stability on the peptides they are linked to and prevent them from premature degradation by endopeptidases. Much attention has been focused on the neurophysiological properties of PGPEPI [3], because it is expressed in the synaptic cleft, where it is involved in the regulation of neuropeptides such as thyrotrophin-releasing hormone, luteinizing hormone (LH)-releasing hormone, and neurotensin [1,2]. However, PGPEPI is widely distributed in various tissues and its actual biological function may depend largely on the type of tissue in which it resides. In the small intestine, which is exposed to dietary xenobiotic peptides, this function may involve the removal of terminal L-pGlu residues to facilitate subsequent degradation by endopeptidases, thereby neutralizing any intrinsic toxic or immunogenic potential. In line with this, it has been proposed that indigestible gluten peptides may evoke coeliac disease (CD) in genetically predisposed individuals [4]. CD is a chronic inflammatory condition of the small intestine, driven by a T helper type 1-adaptive immune response towards peptides released from wheat gluten (and similar storage proteins found in rye and barley). The only possible treatment is adherence to a life-long, gluten-free diet [5,6]. It has recently been demonstrated that some gluten peptides can spontaneously form L-pyroglutamate *in vitro* [7]. If this process also takes place *in vivo* because of insufficient processing by PGPEPI enzyme activity, these peptides may remain resistant to subsequent endopeptidase cleavage, thereby retaining immunogenic properties that are so detrimental to CD patients. We therefore hypothesized that an impaired PGPEPI function may contribute to the risk of developing CD by raising the load of intact gluten antigens resistant to further proteolytic cleavage. One of the enzymes involved in this downstream processing of gluten is prolyl endopeptidase (PREP). Interestingly, the chromosomal locations of both *PGPEPI* (19p13.11) and *PREP* (6q21) coincide with known susceptibility loci for CD [8]. We have performed a detailed functional genomics analysis of *PREP* and ruled out a primary involvement in CD [9]. Here we describe a comprehensive functional genomics

Table 1 **Data on individuals (coeliac disease patients and controls) included in the study**

Patient ID	Age (years)	Sex	Affection status	Atrophy	Diet	DNA for sequencing	Biopsies for activity	Biopsies for expression
1	2	F	CD patient	Atrophy	None	no	yes	no
2	1.5	F	CD patient	Atrophy	None	yes	yes	no
3	1.5	F	CD patient	Atrophy	None	yes	yes	no
4	3.5	M	CD patient	N	GFD	no	yes	no
5	7.5	M	CD patient	Atrophy	GFD	yes	yes	no
6	3	F	CD patient	Atrophy	GFD	yes	yes	no
7	6.5	F	CD patient	Atrophy	None	no	yes	no
8	9	M	CD patient	Atrophy	None	yes	yes	no
9	11	F	CD patient	N	GFD	no	yes	no
10	7	F	CD patient	Atrophy	GFD	yes	yes	no
11	3	M	CD patient	Atrophy	None	yes	yes	no
12	7	F	CD patient	Atrophy	None	yes	yes	no
13	14	F	CD patient	N	GFD	no	yes	no
14	11	F	CD patient	Atrophy	GFD	yes	yes	no
15	14.5	F	CD patient	Atrophy	None	yes	yes	no
16	16.5	F	CD patient	N	GFD	yes	yes	no
17	6	M	CD patient	Atrophy	None	yes	yes	no
18	9.5	F	CD patient	Atrophy	None	yes	yes	no
19	2	F	CD patient	Atrophy	None	yes	yes	no
20	3	M	CD patient	Atrophy	None	yes	yes	no
21	8.5	F	CD patient	N	GFD	no	yes	no
22	14.5	F	CD patient	Atrophy	None	no	yes	no
23	5.5	F	CD patient	Atrophy	GFD	yes	yes	no
24	6	F	CD patient	Atrophy	GFD	yes	yes	no
25	2.5	F	CD patient	Atrophy	None	no	yes	no
26	4.5	F	CD patient	Atrophy	None	no	yes	no
27	16	F	Control	N	None	yes	yes	no
28	12.5	M	Control	N	None	yes	yes	no
29	14.5	M	Control	N	None	yes	yes	no
30	8.5	M	Control	N	None	yes	yes	no
31	3.5	F	Control	N	None	yes	yes	no
32	3	F	Control	N	None	no	yes	no
33	5	F	Control	N	None	yes	yes	no
34	1	M	Control	N	None	no	yes	no
35	4	M	Control	N	None	yes	yes	no
36	15.5	F	Control	N	None	yes	yes	no
37	8.5	M	Control	N	None	yes	yes	no
38	4	M	Control	N	None	yes	yes	no
39	16.5	M	Control	N	None	yes	yes	no
40	2	M	Control	N	None	yes	yes	no
41	12.5	M	Control	N	None	yes	yes	no
42	4	M	Control	N	None	yes	yes	no
43	7.5	M	Control	N	None	yes	yes	no
44	4.5	F	Control	N	None	yes	yes	no
45	10	F	Control	N	None	no	yes	no
46	8.5	F	Control	N	None	no	yes	no
47	8.5	F	Control	N	None	no	yes	no
48	70	F	CD patient	N	GFD	no	no	yes
49	52	F	CD patient	N	GFD	no	no	yes
50	28	F	CD patient	N	GFD	no	no	yes
51	50	M	CD patient	N	GFD	no	no	yes
52	35	F	CD patient	N	GFD	no	no	yes
53	64	M	CD patient	N	GFD	no	no	yes
54	42	F	CD patient	N	GFD	no	no	yes
55	72	F	CD patient	N	GFD	no	no	yes
56	83	F	CD patient	N	GFD	no	no	yes
57	55	F	CD patient	N	GFD	no	no	yes
58	27	M	CD patient	N	GFD	no	no	yes
59	44	M	CD patient	N	GFD	no	no	yes
60	82	F	CD patient	N	GFD	no	no	yes
61	35	F	CD patient	N	GFD	no	no	yes
62	60	F	CD patient	N	GFD	no	no	yes
63	2	F	CD patient	N	GFD	no	no	yes
64	16	F	CD patient	Atrophy	None	no	no	yes
65	2	M	CD patient	Atrophy	None	no	no	yes
66	39	M	CD patient	Atrophy	GFD	no	no	yes
67	57	M	CD patient	Atrophy	None	no	no	yes
68	60	F	CD patient	Atrophy	GFD	no	no	yes
69	55	F	CD patient	Atrophy	GFD	no	no	yes
70	33	F	CD patient	Atrophy	None	no	no	yes
71	2	F	CD patient	Atrophy	None	no	no	yes
72	79	F	CD patient	Atrophy	None	no	no	yes
73	2	M	CD patient	Atrophy	None	no	no	yes
74	19	F	CD patient	Atrophy	None	no	no	yes

Table 1 (continued)

Patient ID	Age (years)	Sex	Affection status	Atrophy	Diet	DNA for sequencing	Biopsies for activity	Biopsies for expression
75	17	F	CD patient	Atrophy	None	no	no	yes
76	62	F	CD patient	Atrophy	None	no	no	yes
77	39	F	Control	N	None	no	no	yes
78	34	F	Control	N	None	no	no	yes
79	32	F	Control	N	None	no	no	yes
80	59	F	Control	N	None	no	no	yes
81	49	F	Control	N	None	no	no	yes
82	49	M	Control	N	None	no	no	yes
83	53	F	Control	N	None	no	no	yes
84	36	F	Control	N	None	no	no	yes
85	74	M	Control	N	None	no	no	yes
86	44	M	Control	N	None	no	no	yes
87	24	F	Control	N	None	no	no	yes
88	36	F	Control	N	None	no	no	yes
89	13	F	Control	N	None	no	no	yes
90	20	F	Control	N	None	no	no	yes
91	28	F	Control	N	None	no	no	yes
92	26	F	Control	N	None	no	no	yes

CD, Coeliac disease; GFD, gluten-free diet; MIII, Marsh stage III; N, normal mucosa.

analysis of *PGPEP1* in relation to CD, which involved DNA sequencing, genetic association testing, quantifying RNA expression, and determining enzyme activity. Based on our results, we conclude that, despite being an attractive functional and positional candidate, the *PGPEP1* gene is not involved in the aetiology or pathology of CD.

Material and methods

Samples

The genetic study was conducted on a cohort of 311 unrelated CD patients and 180 unrelated controls, all of Dutch Caucasian origin. Only CD patients diagnosed according to revised ESPGHAN criteria and with a Marsh III lesion confirmed by duodenal biopsy sampling were selected for this study, as described previously by Van Belzen *et al.* [8] and Walker-Smith *et al.* [10]. Marsh III biopsies are characterized by villous atrophy, crypt hyperplasia, and lymphocytosis. Biopsies from patients in complete remission on a gluten-free diet with normal histology are classified as Marsh 0 [11]. The average age at diagnosis was 34 years (ranging from 1 to 83 years of age). The percentage of females was 69% in the case group and 64% in the control group. The majority of patients were HLA-DQ2 positive (93%). The remaining 7% DQ2-negative patients were either DQ8 positive (50%) or did not carry either DQ2 or DQ8 (50%). We collected blood samples and isolated DNA according to standard laboratory procedures. We also collected biopsy samples from 47 individuals (26 CD patients and 21 controls) for enzyme activity studies (see Table 1, samples 1–47). DNA material was available for sequence analysis from 33 individuals in this group (17 CD patients and 16 controls, who have a normal small bowel; Table 1). We extracted RNA for expression studies from a second set of biopsy samples from 45 individuals (13 untreated MIII patients, 16 M0 patients in complete remission on a gluten-free diet, and 16 controls; Table 1, patients 48–92). The study was approved by the Medical Ethics Committee of the University Medical Centre Utrecht, and informed consent was obtained from the participants.

Sequence analysis

The entire coding sequence of the *PGPEP1* gene was amplified, including the exon–intron boundaries (see Table 2 for primers and protocols). The polymerase chain reaction (PCR) products were examined on a 2% agarose gel and purified using the Millipore Vacuum Manifold, according to the manufacturer's protocol (Billerica, Massachusetts, USA). Samples were prepared using the ABI PRISM BigDye terminator cycle sequencing ready kit (Applied Biosystems, Foster City, California, USA) using the manufacturer's instructions. PCR and sequencing amplification was performed on a GeneAmp PCR system 9700 (Perkin Elmer, Foster City, California, USA). Reaction products were run on a 3730 DNA Analyzer (Applied Biosystems). Sequence analysis and alignment was carried out using the Sequence Navigator (Applied Biosystems) and Vector NTI (InforMax Inc., Massachusetts, USA) software packages.

Genetic association studies and data analysis

Three single-nucleotide polymorphisms (SNP), rs10424419 (assay number: C__25752620_10), rs10421536 (assay number: C__2036245_10) and hCV2036269 (assay number: C__2036269_10), were selected from the *PGPEP1* gene, and the corresponding assays were obtained from Applied Biosystems. These SNP were tested in a case–control study (311 cases and 180 controls) and analysed on an ABI PRISM 7900 HT system (Applied Biosystems). Hardy–Weinberg equilibrium was evaluated separately for cases and controls for all the SNP tested (data not shown). Differences in allele

Table 2 **Sequences of oligonucleotide primers and polymerase chain reaction conditions for *PGPEP1***

	5′ primer	3′ primer	Input genomic DNA (ng)	Protocol
Exon 1	cacgccgagcggctgac	gatttggaaggtgtcatcacc	100	1[a]
Exon 2	cagggtgctcccagaactcag	ggaccttcttggatgccata	50	2[b]
Exon 3	accccagggaccatagaatc	gaagggtgccaacctgtaag	100	3[c]
Exon 4	cccaggcagagctcattaag	Cgcttggcttgttctacacc	50	2
Exon 5	cttcccttggatttggatgg	gatccccaagctgaagagc	50	2

[a]Protocol 1. The reaction was carried out in 25 µl volume containing 100 ng of each forward and reverse primer, 2.5 µl of Pol buffer, 2.5 µl dimethyl sulphoxide, 1.5 µl deoxyribonucleotide triphosphates (dNTP; 25 mM), 0.375 µl of bovine serum albumin and 0.1 units of Taq-polymerase. The cycling consisted of denaturation at 94°C for 4 min, followed by 33 cycles of 94°C for 1 min, 63°C for 1 min, and 72°C for 2 min, with a final extension step at 72°C for 7 min.
[b]Protocol 2. The reactions were carried out in 20 µl volume containing 2 µl dNTP (0.5 mM), 2.5 units of a homemade Taq polymerase, 1.20 µl of 25 mmol magnesium chloride ($MgCl_2$), 50 ng of each oligonucleotide primer, and 2 µl × 10 polymerase chain reaction (PCR) buffer. The cycling consisted of denaturation at 94°C for 7 min, followed by 32 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s. The final extension step was at 72°C for 15 min.
[c]Protocol 3. The PCRs were performed in 20 µl volume containing 2 µl dNTP (0.5 mmol), 1.20 µl of 25 mmol $MgCl_2$, 50 ng of each oligonucleotide primer, 0.8 units of AmpliTaq Gold, and 2 µl PCR Taq Gold Buffer II (Perkin Elmer, Applied Biosystems). The cycling consisted of denaturation at 94°C for 7 min, followed by 32 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s. The final extension step was at 72°C for 7 min.

frequencies and genotype distributions were compared between cases and controls using the χ^2 test.

Quantitative reverse-transcription polymerase chain reaction

Quantification of *PGPEP1* transcripts was performed by quantitative reverse-transcription (qRT) PCR on total RNA as previously described [12]. Commercially available assays were used to measure transcriptional activity of the *PGPEP1* gene (Hs00214546_m1) and that of the endogenous reference gene *GUSB* (PARD 4326320E), both obtained from Applied Biosystems. We synthesized complementary DNA separately from the 16 controls, aliquoted equal amounts in a pool, and used this to perform qRT–PCR. The resulting values were used to normalize the data from the individual cases tested. Both genes were tested in duplicate for all the 29 individual CD patient samples and the control pool on an ABI 7900 HT (Applied Biosystems).

Determination of PGPEPI enzyme activity

To measure PGPEPI catalytic activity, we modified the method described by Dando *et al.* [2]. The duodenal biopsy samples were washed with phosphate-buffered saline, frozen quickly and stored at −80°C for no longer than 18 months. The biopsy samples were thawed on ice and ground using an Ultra Turrax homogenizer (Ika Labortechnik, Staufen, Germany) at 22 000 rpm in the presence of 500 µl of lysis buffer [20 mmol Tris/HCl pH 7.4, 137 mmol sodium chloride, 2 mM ethylenediamine tetraacetic acid (EDTA), 10% glycerol, 1% Triton X-100]. The lysates were centrifuged (14 000 rpm, 15 min, 4°C) and the assay was performed in 96-well black plates with a clear bottom (Corning Inc., New York, USA). All measurements were performed in quadruplicate. Lysates (25 µl) were pre-incubated with 25 µl incubation buffer (100 mmol tripotassium phosphate pH 7.5, 1 mmol EDTA, 1 mmol dithiothreitol) for 5 min at 37°C. The reaction was started by adding 50 µl of substrate solution (20 mmol pGlu-AMC, diluted 50 times in the incubation buffer). The final substrate concentration was 200 µmol. After 2 h incubation at 37°C, the reaction was stopped with 50 µl of 1 mol acetic acid. The concentration of the released 7-amino-4-methylcoumarin (AMC) was measured fluorimetrically at λ_{ex} 360 nm and λ_{em} 460 nm using a CytoFluor multiwell plate reader (PerSeptive Biosciences, Framingham, Massachusetts, USA). One unit of enzyme was defined as the catalytic activity that releases 1 µmol AMC per minute at 37°C. Both pGlu-AMC substrate and standard AMC were purchased from Fluka Chemie AG (Buchs, Switzerland). The total protein concentration in lysates was determined using a Bradford protein assay (Bio Rad, Munich, Germany) with bovine serum albumin (Pierce, Rockford, Illinois, USA) as the standard.

Results

The *PGPEP1* gene is located in the chromosome 19p13 linkage region (CELIAC4 locus), which has an estimated relative risk of 2.6 for CD in the Dutch population [8].

Sequence analysis

PGPEP1 consists of five exons, of which exons 1–4 and part of exon 5 are translated into a protein containing 209 amino acids with a peptidase_C15 domain and a catalytic triad of three amino acids that is highly conserved (Fig. 1). We performed DNA sequence analysis of *PGPEP1* for coding parts of the five exons, including exon–intron boundaries, to search for causal genetic variants that could influence the expression or enzyme activity of PGPEPI. Thirty-seven DNA samples (20 CD patients and 17 normal controls) were used and revealed three polymorphisms: one was novel (basepair position 18.312.567 on chromosome 19, basepair change C/T) and two had previously been reported (Celera names hCV25754037 and hCV2036259; Fig. 1). We observed the new SNP 41 basepairs (bp) downstream of exon 1, in only one control individual. In addition, we confirmed the two other SNP: hCV25754037 was located 40 bp upstream of exon 2, and hCV2036259 was located 18 bp upstream of exon 3 (see Table 3 for genotypes). Based on the Shapiro and Senapathy Splice Score [13], these SNP were predicted to have no influence on the splicing efficiency

Fig. 1

(a)

(b)

	SNP ID	Location (ingene)	Location (bp)	SNP type allele 1/allele 2	Observed minor allele frequency	Used in association study
1	None	intron 1	18312567	C/T	1.6	no
2	rs10424419	intron 1	18314471	G/C	25.0	yes
3	hCV25754037	intron 1	18314629	G/A	25.8	no
4	rs10421536	intron 2	18317963	G/C	9.4	yes
5	hCV2036259	intron 3	18327840	T/C	36.7	no
6	hCV2036269	3' of PGPEP1	18339389	T/A	15.6	yes

The *PGPEP1* gene characterized in the Dutch coeliac disease population. (a) Overview of *PGPEP1* (located on chromosome 19, 18312450–18335428 bps), with untranslated parts of exon 1, 5 and 6 in grey, and translated parts of the exons in black. Numbers give the location of the single-nucleotide polymorphisms (SNP); asterisk gives the location of the microsatellite marker. The three peptidase C15 domains and the location of the expression assay (hCG37051) are shown. (b) Table of the SNP used in this study. SNP 1, 3, and 5 were observed during sequencing of the exons and exon–intron boundaries. SNP 2, 4 and 6 were used for the case–control study. The observed minor allele frequency of SNP 1, 3, 5 is that found in the individuals used for sequencing (either through sequencing or by Taqman typing). The residues of the catalytic triad (GLU85, Cys149 and His169) are located in exons 4 and 5.

of the downstream exons, nor did they introduce new splice or branch sites. We therefore concluded that these SNP are not likely to affect the gene function or play a role in CD susceptibility.

Genetic association studies

Sequencing did not reveal any mutations or SNP that looked likely to have an influence on gene expression or enzyme activity. However, this does not exclude the existence of intronic variants that are associated with the disease. We selected three intronic SNP for genetic association studies in order to define whether genetic variants in *PGPEP1* were involved in CD in the Dutch population. There was hardly any linkage disequilibrium between SNP in *PGPEP1*, so no tag SNP could be defined and the SNP were selected randomly; consequently, this approach could not cover all the genetic variants in the gene. These association studies did not reveal any significant association between *PGPEP1* and CD for the Dutch population (Table 4). We also tested a microsatellite, D19S898, located in intron 4, on 216 cases, 216 controls, and 123 trio families (parents + affected child). This showed no association, although a slight increase in transmission was seen for one allele ($P = 0.094$), not corrected for multiple testing (data not shown).

Quantitative reverse-transcription polymerase chain reaction expression study

We performed *PGPEP1* gene expression studies by qRT–PCR on RNA from duodenal biopsies from 16 CD patients in complete remission on a gluten-free diet with a normalized histology (Marsh 0) and pooled cDNA derived from 16 controls. As the biopsies from the patients and the controls have the same histology, we reasoned that any differences in *PGPEP1* expression might point to an early pathological or even aetiological event in CD, but we observed no differences between these two groups (Fig. 2). We also compared Marsh III (atrophic villi) biopsies of patients to control biopsies, to investigate whether *PGPEP1* might play a secondary role in CD, but again no differences were found. The expression of *PGPEP1* is high compared with *GUSB*, showing that PGPEPI is abundantly present in the intestine. We can conclude from these expression results that any involvement of the *PGPEP1* gene in CD does not arise from differences on the RNA level.

PGPEPI enzyme activity in biopsy samples

The previous results could have missed intronic mutations, as well as association or transcription changes potentially influencing enzyme activity. For example, spurious genetic alterations or post-translational modification can alter enzyme activity. In order to investigate whether an impaired enzymatic activity of PGPEPI has an influence on CD by a hypothetical contribution to the decreased digestion of gluten peptides in the small intestine of CD patients, and thus activates an aberrant immune response, we measured the catalytic activity of the enzyme in 47 biopsies (26 CD patients with varying Marsh stages, of which five were Marsh 0, and 21 controls). The average measured catalytic activity was

Table 3 Activity determinations and genotype data of *PGPEP1*

Patient ID	Affection status	Atrophy	Diet	Activity (μU/mg)	41 bp after exon 1	rs10424419	hCV25754037	rs10421536	hCV2036259
1	CD patient	Atrophy	None	220.79	ND	ND	ND	ND	ND
2	CD patient	Atrophy	None	214.56	11	12	12	22	11
3	CD patient	Atrophy	None	255.99	11	11	11	11	00
4	CD patient	N	GFD	350.38	ND	ND	ND	ND	ND
5	CD patient	Atrophy	GFD	133.66	11	12	12	12	12
6	CD patient	Atrophy	GFD	420.37	11	12	12	22	11
7	CD patient	Atrophy	None	507.50	ND	ND	ND	ND	ND
8	CD patient	Atrophy	None	139.71	11	12	12	22	12
9	CD patient	N	GFD	417.49	ND	ND	ND	ND	ND
10	CD patient	Atrophy	GFD	199.20	11	11	11	22	00
11	CD patient	Atrophy	None	109.02	11	11	11	22	11
12	CD patient	Atrophy	None	206.01	11	12	12	22	11
13	CD patient	N	GFD	348.25	ND	ND	ND	ND	ND
14	CD patient	Atrophy	GFD	416.22	00	11	11	12	12
15	CD patient	Atrophy	None	401.49	11	12	12	22	11
16	CD patient	N	GFD	326.94	11	12	12	22	22
17	CD patient	Atrophy	None	182.64	11	11	11	22	11
18	CD patient	Atrophy	None	9.33	11	11	11	22	11
19	CD patient	Atrophy	None	105.42	11	11	11	12	12
20	CD patient	Atrophy	None	201.12	11	12	12	22	22
21	CD patient	N	GFD	138.58	ND	ND	ND	ND	ND
22	CD patient	Atrophy	None	447.79	ND	ND	ND	ND	ND
23	CD patient	Atrophy	GFD	373.29	11	12	12	22	11
24	CD patient	Atrophy	GFD	384.60	11	12	12	12	22
25	CD patient	Atrophy	None	123.78	ND	ND	ND	ND	ND
26	CD patient	Atrophy	None	103.77	ND	ND	ND	ND	ND
27	Control	N	None	421.58	00	11	11	22	11
28	Control	N	None	316.58	11	11	12	22	22
29	Control	N	None	213.75	11	11	11	22	12
30	Control	N	None	466.12	11	12	12	22	12
31	Control	N	None	278.61	11	11	11	22	12
32	Control	N	None	414.03	ND	ND	ND	ND	ND
33	Control	N	None	198.37	11	11	11	22	11
34	Control	N	None	321.25	ND	ND	ND	ND	ND
35	Control	N	None	259.89	11	12	12	22	12
36	Control	N	None	204.02	11	12	12	22	12
37	Control	N	None	201.91	11	11	11	22	11
38	Control	N	None	331.37	11	22	22	22	12
39	Control	N	None	480.26	12	11	11	22	22
40	Control	N	None	76.51	11	11	11	22	11
41	Control	N	None	228.43	11	00	11	00	00
42	Control	N	None	259.49	11	12	12	22	12
43	Control	N	None	187.38	11	11	11	22	12
44	Control	N	None	37.59	11	11	11	22	11
45	Control	N	None	253.78	ND	ND	ND	ND	ND
46	Control	N	None	271.77	ND	ND	ND	ND	ND
47	Control	N	None	173.69	ND	ND	ND	ND	ND

CD, Coeliac disease; GFD, gluten-free diet; N, normal biopsies from controls; PGPEP1 activity determinations in duodenal biopsies.

Table 4 *P* values obtained from testing the case–control cohort for three intronic single-nucleotide polymorphisms

SNP name	SNP type	No. of cases (MAF%)	No. of controls (MAF%)	*P* value
rs10424419	G/C	305 (26.4)	174 (28.5)	0.492
rs10421536	G/C	307 (7.8)	180 (6.7)	0.508
hCV2036269	T/A	301 (16.9)	180 (13.1)	0.119

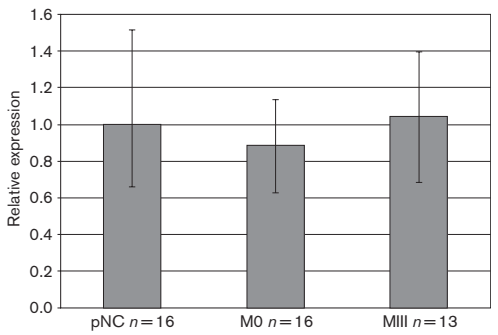
MAF, Minor allele frequency; SNP, single-nucleotide polymorphism.
All the SNP were tested on 311 cases and 180 controls.

0.266 U/g protein, with a large individual variation (from 0.09 to 0.51 U/g protein). We did not observe statistically significant differences in PGPEPI enzyme activity when we compared biopsies from patients with control biopsies, nor when comparing between normalized biopsies from remission patients with control biopsies (Fig. 3, enzyme activity data of all biopsies given in Table 3). Furthermore, we were unable to correlate PGPEPI enzyme activity levels with the histological status, diet, age or sex of the individuals studied (data not shown).

Discussion

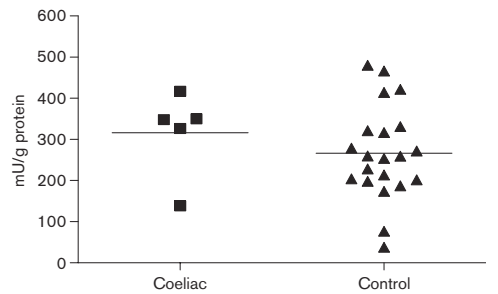
PGPEPI has been studied for its role in neurophysiological processes [3], but its role in the small intestine has never been looked at. Our study is the first to investigate the role of PGPEPI in the small intestine, especially in relation to CD. PGPEPI has been described mainly as a cytosolic enzyme. It is unknown whether PGPEPI is secreted from the pancreas or is present in the brush border of enterocytes, nor if that is required for its

Fig. 2



Expression study for *PGPEP1* using small intestine biopsies. A pool of 16 control biopsies (pNC) was compared with 16 normalized biopsies (Marsh 0) of patients on a gluten-free diet and with 13 Marsh III (MIII) biopsies from newly diagnosed patients. All expression levels were corrected for the amount of RNA. The expression level of the pool of control biopsies was set to 1; the other expression levels are relative to the level of the pool of controls.

Fig. 3



PGPEPI enzyme activity in small intestine biopsies. Biopsies from five coeliac cases with a normalized histology (Marsh 0) were compared with biopsies from normal controls. Unpaired *t* test ($P=0.389$).

function in the small intestine. However, its presence in secretory fluids such as human semen has been reported [14]. The small intestine is one of the first entry points for metabolites into our internal environment, and it acts as the first line of defence to detrimental proteins and organisms by using several enzymes. CD is caused by an immune response to gluten peptides. Gluten is extremely rich in the amino acid glutamine, which, when present at the *N*-terminus of gluten-derived peptides, can be converted into pyroglutamate. The rate of conversion of gluten-derived peptides into pyroglutamyl-containing peptides has not been studied, making it difficult to assess the exact contribution of PGPEPI on the digestibility and detoxification of gluten. In-vitro studies have shown that PGPEPI cleaves such pyroglutamate residues from gluten peptides, which facilitates their proteolysis [7].

We therefore hypothesized that a defect in PGPEPI enzyme activity might play a role in CD development, as this could result in a less efficient degradation of gluten peptides in which an *N*-terminal pyroglutamate residue is present. Investigation of the genetic risk factors in CD has, so far, only resulted in the identification of the HLA-DQ2 and HLA-DQ8 genes and some evidence for a role of *CTLA4* or its neighbouring genes [15,16]. CTLA4 plays a role by regulating T-cell reactivity, and the role of the HLA-DQ genes in CD has been clarified by their ability to present gluten to T cells. Besides this, little is known about the type of genes involved in the genetics of CD. Several genome-wide screens have been performed, which has recently resulted in the discovery of two new linkage regions in the Dutch CD population on chromosome 6q21 and on chromosome 19p13 [8]. In both regions, there are genes that have endopeptidase activity, thereby making them functional and positional candidates for CD. The *PREP* gene on chromosome 6q21 has been studied for its role in CD, but no genetic association was found [9]. Another candidate gene, *PGPEP1*, resides in the linkage region on chromosome 19p13. In this study we searched for mutations or SNP in *PGPEP1*, by sequencing exons and exon–intron boundaries that could influence the gene expression or enzyme activity. We observed no SNP or mutations that could potentially influence PGPEPI function. We also selected three intronic SNP for a case–control study, but we found no association between any of the alleles and CD. We performed gene expression and enzyme activity studies to confirm that we had not missed any causative relationship between *PGPEP1* and CD. We show here that PGPEPI has a relatively high level of expression in the small intestine, which has not been reported before. Although a large group of patients and controls were studied, no significant differences in *PGPEP1* gene expression or enzymatic activity were observed when comparing normalized or Marsh III duodenal biopsies in patients with control biopsies.

In this study we have demonstrated that *PGPEP1*, although being an attractive positional and functional candidate gene, is not a causative factor in CD pathogenesis. Although we excluded both gluten-processing peptidases, *PREP* and *PGPEP1*, as candidate genes for chromosome 6q21 and CELIAC4, respectively, we cannot completely rule out that other peptidases involved in gluten detoxification may contribute to the genetic risk of CD. In parallel to this research, we recently identified *MYO9B* as the CELIAC4 gene in the 19p13

region [17]. The structure of this gene points to a possible causative mechanism of gluten peptides passage through the epithelial barrier. This stresses the involvement of non-immune-related genes and pathways, like gluten-transport and gluten-processing, as potential genetic factors for CD susceptibility.

Acknowledgements

The authors would like to thank J. Senior for editing the text, R. van't Slot for technical assistance, Dr J.W.R. Meijer and Professor G.A. Meijer for reviewing the patients' biopsies, Dr J.J. Schweizer for help in collecting the biopsies and Professor C.J. Mulder and Dr R. Houwen for referring patients to our research project.

Conflict of interest

The authors declare no conflict of interest.

Authors' contributions

A.J.M. performed and analysed the genetic studies and was involved in writing the manuscript; D.S. performed and analysed the enzymatic studies; B.D. performed and analysed the gene expression studies; M.C.W. was involved in writing and editing the manuscript; M.L.M. obtained the materials for the expression and enzymatic studies; F.K. and C.W. had the original idea for the study.

References

1 Cummins PM, O'Connor B. Pyroglutamyl peptidase: an overview of the three known enzymatic forms. *Biochim Biophys Acta* 1998; **1429**:1–17.
2 Dando PM, Fortunato M, Strand GB, Smith TS, Barrett AJ. Pyroglutamyl-peptidase I: cloning, sequencing, and characterisation of the recombinant human enzyme. *Protein Expr Purif* 2003; **28**:111–119.
3 Larrinaga G, Callado LF, Agirregoitia N, Varona A, Gil J. Subcellular distribution of membrane-bound aminopeptidases in the human and rat brain. *Neurosci Lett* 2005; **383**:136–140.
4 Shan L, Molberg O, Parrot I, Hausch F, Filiz F, Gray GM, *et al.* Structural basis for gluten intolerance in coeliac sprue. *Science* 2002; **297**: 2275–2279.
5 Mowat AM. Coeliac disease – a meeting point for genetics, immunology, and protein chemistry. *Lancet* 2003; **361**:1290–1292.
6 Papadopoulos GK, Wijmenga C, Koning F. Interplay between genetics and the environment in the development of coeliac disease: perspectives for a healthy life. *J Clin Invest* 2001; **108**:1261–1266.
7 Higaki-Sato N, Sato K, Esumi Y, Okumura T, Yoshikawa H, Tanaka-Kuwajima C, *et al.* Isolation and identification of indigestible pyroglutamyl peptides in an enzymatic hydrolysate of wheat gluten prepared on an industrial scale. *J Agric Food Chem* 2003; **51**:8–13.
8 Van Belzen MJ, Meijer JW, Sandkuijl LA, Bardoel AF, Mulder CJ, Pearson PL, *et al.* A major non-HLA locus in coeliac disease maps to chromosome 19. *Gastroenterology* 2003; **125**:1032–1041.
9 Diosdado B, Stepniak DT, Monsuur AJ, Franke L, Wapenaar MC, Mearin ML, *et al.* No genetic association of the human prolyl endopeptidase gene found in the Dutch coeliac disease population. *Am J Physiol Gastrointest Liver Physiol* 2005; **289**:G495–G500.
10 Walker-Smith JA, Guandalini S, Schmitz J, Shmerling DH, Visakorpi JK. Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. *Arch Dis Child* 1990; **65**:909–911.
11 Marsh M. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 1992; **102**: 330–354.
12 Wapenaar MC, van Belzen MJ, Fransen JH, Sarasqueta AF, Houwen RH, Meijer JW, *et al.* The interferon gamma gene in coeliac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. *J Autoimmun* 2004; **23**:183–190.
13 Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl Acids Res* 1987; **15**:7155–7174.
14 O'Cuinn G, O'Connor B, Elmore M. Degradation of thyrotropin-releasing hormone and luteinising hormone-releasing hormone by enzymes of brain tissue. *J Neurochem* 1990; **54**:1–13.
15 Hunt KA, McGovern DP, Kumar PJ, Ghosh S, Travis SP, Walters JR, *et al.* A common CTLA4 haplotype associated with coeliac disease. *Eur J Hum Genet* 2005; **13**:440–444.
16 Van Heel DA, Hunt K, Greco L, Wijmenga C. Genetics in coeliac disease. *Best Pract Res Clin Gastroenterol* 2005; **19**:323–339.
17 Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, *et al.* Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. *Nat Genet* 2005; **37**:1341–1344.

Chapter 2.2

No genetic association of the human prolyl endopeptidase gene in the Dutch celiac disease population

Alienke J. Monsuur, Begoña Diosdado, Dariusz T. Stepniak, Lude Franke, Martin C. Wapenaar, Maria Luisa Mearin, Frits Koning, and Cisca Wijmenga

Am J Physiol Gastrointest Liver Physiol 2005; 289: G495–G500

No genetic association of the human prolyl endopeptidase gene in the Dutch celiac disease population

Begoña Diosdado,[1]* Dariusz T. Stepniak,[2]* Alienke J. Monsuur,[1]* Lude Franke,[1] Martin C. Wapenaar,[1] Maria Luisa Mearin,[3] Frits Koning,[2] and Cisca Wijmenga[1]

[1]*Complex Genetics Section, Department of Biomedical Genetics, University Medical Centre, Utrecht, and Departments of* [2]*Immunohematology and Blood Transfusion and* [3]*Paediatrics, Paediatrician Unit of Paediatric Gastroenterology, Leiden University Medical Centre, Leiden, the Netherlands*

Submitted 8 February 2005; accepted in final form 28 April 2005

Diosdado, Begoña, Dariusz T. Stepniak, Alienke J. Monsuur, Lude Franke, Martin C. Wapenaar, Maria Luisa Mearin, Frits Koning, and Cisca Wijmenga. No genetic association of the human prolyl endopeptidase gene in the Dutch celiac disease population. *Am J Physiol Gastrointest Liver Physiol* 289: G495–G500, 2005. First published May 12, 2005; doi:10.1152/ajpgi.00056.2005.—Celiac disease (CD) is a complex genetic disorder of the small intestine. The DQ2/DQ8 human leucocyte antigen (HLA) genes explain ~40% of the genetic component of the disease, but the remaining non-HLA genes have not yet been identified. The key environmental factor known to be involved in the disease is gluten, a major protein present in wheat, barley, and rye. Integrating microarray data and linkage data from chromosome 6q21–22 revealed the prolyl endopeptidase (*PREP*) gene as a potential CD candidate in the Dutch population. Interestingly, this gene encodes for the only enzyme that is able to cleave the proline-rich gluten peptides. To investigate the role of the human *PREP* gene as a primary genetic factor in CD, we conducted gene expression, sequence analysis, and genetic association studies of the *PREP* gene and determined PREP enzyme activity in biopsies from CD patients and controls. Sequence analysis of the coding region of the *PREP* gene revealed two novel polymorphisms. Genetic association studies using two novel polymorphisms and three known *PREP* variants excluded a genetic association between *PREP* and CD. Determination of PREP activity revealed weak but significant differences between treated and untreated CD biopsies ($P < 0.05$). Our results from the association study indicate that *PREP* is not a causative gene for CD in the Dutch population. These are further supported by the activity determinations in which we observed no differences in PREP activity between CD patients and controls.

CELIAC DISEASE (CD) is a chronic autoimmune disorder caused by the ingestion of dietary gluten. Gluten toxicity in CD patients is, in part, determined by the proline- and glutamine-rich gliadins, secalins, and hordeins present in wheat, rye, and barley, respectively. This toxicity results from the presence of a repertoire of T-cells in the lamina propria of the intestines of CD individuals that are able to recognize many different gluten peptides and provoke an erroneous immune response in the small intestine. This leads to specific tissue damage characterized by lymphocytic infiltration of the mucosa (Marsh I), a Marsh II stage presenting crypt hyperplasia together with the Marsh I features, and Marsh III (MIII) stage in which, in addition to Marsh II, villous atrophy develops (4, 13, 15).

So far, the only treatment for CD patients is a strict gluten-free diet, but new alternatives have been recently proposed based on an improved understanding of the disease ethiopathogenesis (11, 12, 14). One of the most attractive new approaches consists of an enzymatic therapy using bacterial prolyl-endopeptidase from *Flavobacterium meningosepticum*, an enzyme that can remove gluten toxicity by cleaving it into small fragments that lack T-cell stimulatory properties (11). This bacterial enzyme has a well-conserved evolutionary homologue in humans (EC 3.4.21.26) (17) that encodes for a cytosolic enzyme that also hydrolyzes amide bonds of very rich proline peptides shorter than 30 amino acids (18). It is tempting to speculate that an impaired function of the human prolyl endopeptidase (PREP) would result in the accumulation of long, immunostimulatory gluten peptides in the lumen or lamina propria, and that this could play a role in breaking down an individual's tolerance to gluten.

Interestingly, the human *PREP* gene is located in the chromosomal region 6q21–22 that showed suggestive linkage (lod score 3.10, $P = 1.3 \times 10^{-4}$) to CD in the Dutch population (16). In addition, microarray experiments performed in the same population showed an approximately twofold upregulation of *PREP* in seven untreated CD patients compared with four treated CD patients, all 11 of whom still showed villous atrophy ($P < 0.005$) (5).

Because these results suggested a role for the human *PREP* gene as a primary candidate for CD in the Dutch population, we performed a detailed analysis of PREP activity and follow-up expression in biopsies of patients and controls, sequenced the *PREP* gene in a large group of patients, and carried out genetic association studies.

MATERIALS AND METHODS

Subjects. Seven CD patients from seven not related sibpairs who contributed to the linkage peak on chromosome 6p21–22 and showed two alleles identical by descent for this region were selected for resequencing the *PREP* gene to define new variants in exon and exon-intron boundaries.

We collected 47 biopsies for the enzyme activity studies (Table 1) from 24 CD patients with an MIII biopsy proven lesion, and 23 controls who had a biopsy examination for other reasons, such as abdominal pain or failure to thrive. The diagnosis of the CD patients was done according to the ESPGHAN criteria (20). DNA material was available for 37 of these samples [18 CD patients (*individuals 24-41*,

* B. Diosdado, D. T. Stepniak, and A. J. Monsuur contributed equally to this work.

Address for reprint requests and other correspondence: C. Wijmenga, Complex Genetics Section, Stratenum 2.117, Dept. of Biomedical Genetics, Univ. Medical Centre Utrecht, PO Box 85060, 3508 AB Utrecht, The Netherlands (e-mail: t.n.wijmenga@med.uu.nl).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Table 1. *Data on individuals (CD patients and controls) included in the study*

Patient	Age, yr	Gender	Status	Number of Biopsies	Histological Stage	Diet	DQ2	DQ8	EMA-IgA	TGA-IgA	Clinical Symptoms
1	4	M	Control		Control	None	Pos	Neg	Neg	Neg	Lassitude
2	17	F	Control		Control	None	ND	ND	Neg	ND	Epigastric pain
3	16	M	Control		Control	None	ND	ND	ND	ND	Diarrhea, abdominal pain
4	15	M	Control		Control	None	ND	ND	ND	ND	Diarrhea
5	10	M	Control		Control	None	Pos	Neg	ND	Neg	Short stature
6	16	M	Control		Control	None	ND	ND	ND	ND	Chronic vomiting
7	9	M	Control		Control	None	Pos	Neg	ND	ND	Constipation
8	2	M	Control		Control	None	ND	ND	Neg	Neg	Failure to thrive
9	5	F	Control		Control	None	ND	ND	ND	ND	Suspected CD
10	6	F	Control		Control	None	Pos	Neg	Neg	Neg	Diarrhea
11	13	M	Control		Control	None	ND	ND	ND	ND	Diarrhea and anal fistels
12	18	M	Control		Control	None	Neg	Neg	Neg	Neg	Abdominal pain
13	4	F	Control		Control	None	ND	ND	ND	ND	Short stature, constipation
14	1	M	Control		Control	None	ND	ND	dubious	ND	Vomits and failure to thrive
15	5	M	Control		Control	None	Neg	Neg	Neg	Neg	Abdominal pain
16	5	M	Control		Control	None	Pos	Neg	Neg	Neg	Abdominal pain
17	14	M	Control		Control	None	ND	ND	ND	ND	Epigastric pain
18	4	M	Control		Control	None	Pos	Pos	Pos	ND	Vomits
19	11	F	Control		Control	None	ND	ND	ND	ND	Unknown
20	4	F	Control		Control	None	Pos	Neg	Neg	ND	Suspected CD
21	10	F	Control		Control	None	ND	ND	Neg	ND	Epigastric pain
22	9	F	Control		Control	None	ND	ND	ND	ND	Suspected CD
23	9	F	Control		Control	None	ND	ND	ND	ND	Suspected CD
24	2	F	CD patient	1st (Diagnostic)	MIII	None	Pos	Neg	Pos	Pos	Asymptomatic
25	15	F	CD patient	1st (Diagnostic)	MIII	None	ND	ND	Pos	ND	Unknown
26	3	F	CD patient	1st (Diagnostic)	MIII	None	Pos	Neg	Pos	Pos	Chronic diarrhea, lassitude
27	10	F	CD patient	1st (Diagnostic)	MIII	None	ND	ND	Pos	ND	Unknown
28	10	M	CD patient	1st (Diagnostic)	MIII	None	Pos	Neg	Pos	ND	Abdominal pain
29	7	M	CD patient	1st (Diagnostic)	MIII	None	Pos	Pos	Pos	Pos	Chronic diarrhea, lassitude
30	6	F	CD patient	1st (Diagnostic)	MIII	None	ND	ND	ND	ND	None
31	3	M	CD patient	1st (Diagnostic)	MIII	None	ND	ND	ND	ND	Unknown
32	7	F	CD patient	3rd (Challenge)	MIII	Challenge	Pos	Neg	Pos	Pos	Asymptomatic
33	9	M	CD patient	2nd (Control)	MIII	GFD	Pos	Neg	ND	ND	Unknown
34	4	F	CD patient	2nd (Control)	M0	GFD	Pos	Neg	ND	ND	None
35	6	F	CD patient	2nd (Control)	MI-II	GFD	Pos	Neg	ND	ND	Unknown
36	17	F	CD patient	2nd (Control)	M0	GFD	Pos	Neg	ND	ND	None
37	9	F	CD patient	2nd (Control)	M0	GFD	Pos	Neg	Neg	Neg	None
38	6	F	CD patient	2nd (Control)	MI-II	GFD	Pos	Neg	ND	ND	Unknown
39	8	F	CD patient	2nd (Control)	MI	GFD	Pos	Neg	Neg	Neg	None
40	2	F	CD patient	2nd (Control)	M0	GFD	Pos	Neg	Neg	Neg	None
41	12	F	CD patient	2nd (Control)	M0	GFD	Pos	Neg	Neg	Neg	None
42	5	F	CD patient	1st (Diagnostic)	MIII	None	ND	ND	ND	ND	Failure to thrive
43	15	F	CD patient	1st (Diagnostic)	MIII	None	ND	ND	Pos	Pos	Chronic diarrhea
44	3	F	CD patient	1st (Diagnostic)	MIII	None	ND	ND	ND	ND	Unknown
45	3	F	CD patient	1st (Diagnostic)	MIII	None	ND	ND	ND	ND	Unknown
46	15	F	CD patient	2nd (Control)	M0	GFD	ND	ND	Neg	ND	None
47	4	M	CD patient	2nd (Control)	M0	GFD	Pos	Neg	Neg	Neg	None
48					M*						
49					M*						
50					M*						
51					M*						
52					M*						
53					M*						
54					M*						

M*, independent sibpairs from genome screen; M, male; F, female; EMA-IgA, antibodies anti-endomysium; TGA-IgA, antibodies anti-gliadin; CD, celiac disease patients; GFD, gluten-free diet; Neg, negative; Pos, positive; M0, CD patients in complete remission; MIII, Marsh III stage; ND, not determined.

Table 1) and 19 controls (*individuals 1-19*, Table 1)], which allowed us to assess both genotype and activity data.

The genetic study comprised a group of 311 independent CD cases and 180 independent age- and sex-matched random hospital controls, all of Dutch Caucasian origin. Only CD patients with a biopsy-proven MIII lesion were included in this study. We collected blood samples and isolated DNA according to standard laboratory procedures (16).

Initially, 16 biopsies from eight MIII CD patients and eight M0 CD patients and a pool of 16 RNA samples from control individuals were used to validate the microarray results for the *PREP* gene using real-time RT-PCR (see Table 1 of supplemental data at http://ajpgi.physiology.org/cgi/content/full/00056.2005/DC1). These samples were not used in the further studies.

The study was approved by the Medical Ethics Committees of the University Medical Centres in Utrecht and Leiden, and informed consent was obtained from all individuals.

Determination of PREP enzyme activity. To measure the PREP activity, we modified the method described by Goossens et al. (8). The duodenal biopsies were washed with PBS, frozen, and stored at

−80°C for no longer than 18 mo. The biopsies were thawed on ice and ground with an Ultra Turrax homogenizer (Ika Labortechnik, Staufen, Germany) at 22,000 rpm in the presence of 500 μl of lysis buffer (20 mM Tris·HCl, pH 7.4, 137 mM NaCl, 2 mM EDTA, 10% glycerol, and 1% Triton X-100). The lysates were centrifuged (14,000 rpm, 15 min, 4°C) and the assay was performed in 96-well black plates with a clear bottom (Corning). Every measurement was performed four times. Twenty microliters of lysates were preincubated with 75 μl of incubation buffer (100 mM K_3PO_4, pH 7.5, 1 mM EDTA, and 1 mM DTT) for 5 min at 37°C. The reaction was started by adding 5 μl of substrate solution [4 mM Z-Gly-Pro-7-amino-4-methylcoumarin (pro-AMC) in 60% methanol]. After 1 h of incubation at 37°C, the reaction was stopped with 50 μl of 1 M acetic acid. The concentration of the released AMC was measured fluorimetrically at an excitation wavelength of 360 nm and an emission wavelength of 460 nm using a CytoFluor multi-well plate reader (PerSeptive Biosciences). One unit of the enzyme was defined as the catalytic activity that releases 1 μmol of AMC per minute. Both Z-Gly-Pro-AMC substrate and standard AMC were purchased from Fluka Chemie (Buchs, Switzerland). Total protein concentration in lysates was determined using a Bradford protein assay (Bio-Rad, Munchen, Germany) and a BCA protein assay (Pierce, Rockford, IL), with BSA (Pierce) as the standard in both cases.

Quantitative real-time RT-PCR. Quantification of *PREP* transcriptional activity was performed by real-time RT-PCR on RNA from biopsies as previously described (19). We used an Assay-on-Demand Gene Expression product for the *PREP* gene (ABI Hs.00267576), and the *GUSB* gene (detected by PARD 4326320E) as an endogenous reference to correct for expression-independent sample-to-sample variability (Applied Biosystems, Foster City, CA). To quantify the relative expression by the $2^{-\Delta\Delta Ct}$ method (19), equimolar amounts of total RNA from 16 control individuals were pooled and used for normalization of the expression data. Both genes were tested in duplicate for all the individual patient samples and the control pool on an ABI 7900 HT (Applied Biosystems).

Sequence analysis. PCR amplification was performed on all 15 exons and exon-intron boundaries of the *PREP* gene. Details about the primer sequences and the PCR conditions can be found in Table 2 of the supplementary data. The PCR products were examined on a 2% agarose gel and purified with the Millipore Vacuum Manifold (Billerica, MA), according to the manufacturer's protocol. Samples were prepared with the ABI PRISM BigDye terminator cycle sequencing ready kit (Applied Biosystems) according to the manufacturer's protocol. PCR and sequencing amplification were performed on a GeneAmp PCR system 9700 (Perkin-Elmer, Wellesley, MA). Sequencing was performed on a 3730 DNA sequencer (Applied Biosystems). Analysis and alignment was carried out with the Sequence Navigator (Applied Biosystems) and Vector NTI (InforMax).

Genetic association studies and data analysis. Five of the selected single nucleotide polymorphism (SNPs) were typed using assay-on-demand probes from the *PREP* gene: hCV1963751 (ABI no. C___1963751_10), rs9486069 (ABI no. C__11638424_10), rs1078725 (ABI no. C___8304693_10), rs2793389 (ABI no. C__11635753_10), and rs1051484 (ABI no. C___8304751_20). The sixth selected SNP, rs12192054, was typed by using an assay-by-design probe from Applied Biosystems. These SNPs were tested in a case-control study (311 cases and 180 controls) and analyzed on an ABI Prism 7900 HT system (Applied Biosystems). Hardy-Weinberg equilibrium was evaluated separately in cases and control, for all SNPs tested (data not shown). Differences in allele frequencies and genotype distributions were compared between cases and controls using the χ^2-test.

RESULTS

We compiled our earlier microarray (5) and linkage (16) data from Dutch CD patients using TEAM, a bioinformatics tool developed in house (6), which allowed us to define the physical location of the differentially expressed genes under the genetic linkage peaks. Integration and analysis of these two data sets revealed that *PREP* was one of the differentially expressed genes located under the linkage peak on chromosome 6q21–22 in the Dutch genome screen (Fig. 1, *A* and *B*). The 6q21–22 region encompasses 22 megabases and contains 111 genes. The relative risk in the Dutch CD population attributed to this locus is 2.3 (16). Quantitative expression studies by real-time RT-PCR on a set of eight RNA samples from treated CD patients in complete remission (M0), eight untreated CD patients with total villus atrophy (MIII), and a pool of normal controls validated these findings. The experiments showed that *PREP* was significantly downregulated in treated M0 patients compared with MIII patients ingesting gluten (1.3-fold, $P < 0.05$; Table 1 of supplementary data), although to a lesser extent than previously described (5).

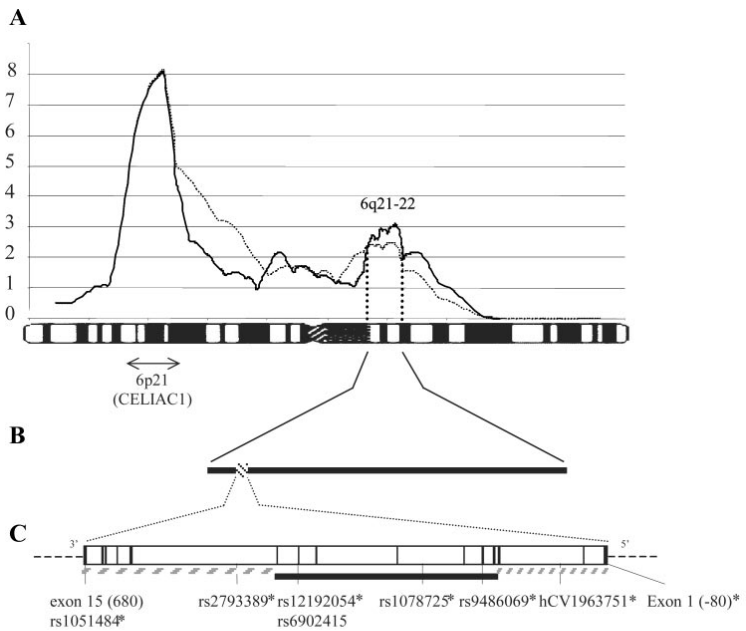
Sequence analysis. To investigate whether the enzymatic properties of this gene product or its expression levels were different in CD patients due to an underlying genetic variation, we performed sequence analysis on the entire coding region and exon-intron boundaries of the *PREP* gene to identify putative mutations or variants in the CD population. The human *PREP* gene is fully annotated in the public databases and consists of 2,905 nucleotides distributed over 15 exons that encode 710 amino acid residues (17). Because the tertiary structure of the human PREP has not yet been described, we used the tertiary structure of its porcine homologue as reference in defining which exons were encoded by which domains. The human and porcine enzymes are 97% homologous at the amino acid level, and in the porcine PREP, exons 1–3 and 10–15 encode the catalytic domain and exons 3–10, the characteristic β-propeller domain that regulates its proteolitic activity (Fig. 1*C*) (7).

Sequence analysis of all 15 exons and exon-intron boundaries in 44 individuals revealed six SNPs in the coding region of *PREP*. These SNPs were present in exon 1, exon 5, exon 9 (two SNPs), and exon 15 (two SNPs) (Fig. 1*B*). The SNP in exon 1 and one of the two SNPs in exon 15 have not yet been annotated in public databases. The published allele frequencies and the frequency of occurrence of these SNPs in the sequenced individuals are shown in Fig. 1*D*.

Only two of the identified SNPs lead to an amino acid change in the PREP protein. A SNP found in exon 9, 1050T→G, gives rise to a leucine to valine substitution at position 351 while a SNP in exon 15, 2118 G→A, gives rise to a valine to isoleucine substitution at position 706. This latter substitution is not expected to have any impact on the function of the PREP protein because the amino acid at position 706 is not conserved (valine in man, isoleucine in pigs, bovines, rats and mice). The leucine-to-valine substitution at position 351 is a conservative one and, therefore, we cannot rule out that this substitution may impact PREP function.

Genetic association studies. To further investigate whether genetic polymorphisms in PREP are associated with CD in the Dutch population, we performed genetic association studies. For our linkage peak on chromosome 6p22, with a relative risk of 2.3 and a SNP frequency in the range of 0.1–0.4, our sample size had 80% power to detect a confidence interval of 95%.

Four exonic SNPs [exon 1 (−80), rs9486069, rs12192054, and rs1051484] were selected based on their high heterozygos-



D

SNP ID	Exon location	Base pair location	Amino acid position	Nucleotide change	Amino acid change	Allele frequency in public databases	Found in 44 sequenced individuals
exon1 (-80)*	exon 1	105896438	(-80)	**G**/**T**	UTR 5'	Not present	33 G/G, 11 G/T
hCV1963751*	intron 2	105891476	-	G/A	-	0.66G/0.34A	ND
rs9486069*	exon 5	105867066	130	TA**T**/TA**C**	Tyr/Tyr	0.66T/0.34C	17 T/T, 22 C/T, 1 C/C 4 ND
rs1078725*	intron 6	105858273	-	T/C	-	0.77C/0.23T	ND
rs12192054*	exon 9	105822511	351	**T**TA/**G**TA	Leu/Val	ND	29 T/T, 9 T/G, 2 G/G, 4 ND
rs6902415	exon 9	105822399	375	TT**C**/TT**T**	Phe/Phe	ND	43 C/C, 1C/T
rs2793389*	intron 10	105816736	-	C/A	-	0.85C/0.15A	ND
exon15 (680)	exon 15	105771761	680	CA**C**/CA**T**	His/His	Not present	43 C/C, 1T/T
rs1051484*	exon 15	105771702	706	**G**TC/**A**TC	Val/Ile	0.72G/0.28A	33 G/G, 11 G/A

Fig. 1. *A*: linkage data of 101 sibpairs [Dutch celiac disease (CD) patients] on chromosome 6. The dashed line indicates the linkage graph before fine mapping, whereas the continuous line is after fine mapping. *B*: 95% confidence interval containing 111 genes; the dashed square indicates the position of the prolyl endopeptide (*PREP*) gene. *C*: exonic-intronic view of *PREP*. The dashed line represents the catalytic domains of the protein and the continuous line the β-propeller domain. *D*: table includes the 6 exonic single nucleotide polymorphisms (SNPs) identified by sequencing in 44 individuals and the 3 intronic SNPs. *SNPs selected for the genetic studies. SNP ID, SNP number; UTR, untranslated region; ND, not determined.

ity in our sequence samples and their possible influence on the protein. Unfortunately, the SNP in the 5′-untranslated region (5′-UTR) could not be designed because of the extreme repetitiveness in the region. None of the three SNPs, however, showed a statistical difference between the cases and controls (Table 2).

To further exclude *PREP* as a causative gene, we selected three noncoding SNPs for further genetic association studies on the basis of a minor allele frequency of >10% (Table 2). These three SNPs also showed a lack of statistical difference between the cases and controls (Table 2). Haplotype analysis did not change these results (data not shown). Finally, we tested a microsatellite marker located in intron 2 of the *PREP* gene, which also showed no association with CD (data not shown). Overall, we found no association between any of our genetic markers and CD.

Table 2. *P values obtained from testing the case-control cohort for three coding and three intronic SNPs*

SNP Name	SNP Type	MAF*	Number of Cases of MAF	Number of Controls	
				MAF	*P* value
hCV1963751	G/A	0.34 (A)	308 (34.7)	178 (36.5)	0.577
rs9486069	T/C	0.34 (C)	306 (28.3)	175 (31.4)	0.301
rs1078725	T/C	0.23 (T)	309 (22.0)	177 (20.9)	0.688
rs12192054	G/T	ND (G)	305 (16.4)	176 (12.2)	0.079
rs2793389	C/A	0.15 (A)	307 (21.5)	179 (18.2)	0.211
rs1051484	A/G	0.28 (A)	309 (15.4)	180 (14.7)	0.784

Values in parentheses are percentages. *Minor allele frequencies (MAF) were obtained from the Celera or Ensemble databases. All single nucleotide polymorphisms (SNPs) were tested on 311 cases and 180 controls.

Activity of PREP in biopsy material from patients and controls. To further investigate whether an impaired enzymatic activity of PREP could be responsible for a decreased digestion of gluten peptides in the small intestine of CD patients and, hence, activation of an aberrant immune response, the catalytic activity of the enzyme was measured in 47 biopsies from CD patients and controls. The activity values lay in the range of 1.71 to 8.52 U/g protein with an average of 4.8 U/g protein (SD = 1.61), which is in agreement with the described PREP activities measured in other human tissues (8). First, patients were grouped according their histological status and adhering to the treatment in treated CD (M0) and untreated CD (MIII) and independently of their genotypes. The average PREP activity levels measured in the untreated CD patients were lower than in the treated CD patients ($P < 0.05$). No significant differences were observed between the treated or untreated CD patients and the controls (Fig. 2). We were not able to correlate PREP activity levels with the age or gender of the studied individuals (data not shown).

Activity-genotype correlations. To further detect an influence of the tested genetic variants on the expression and activity results, we calculated whether there was any association between the different genotypes of the SNPs and the enzymatic activity of PREP.

For activity-genotype correlation, the genotypes of four identified coding SNPs of the gene (Fig. 1*D*) and the activity measurements of 37 individuals were studied (Table 1, *individuals 1–19* and *24–41*). To do so, individuals were grouped according to their genotypes and the average of the activity for each group was calculated for each of the coding SNPs [except SNP rs6902415, because all individuals were homozygote C/C, and exon 15 (680), because all individuals but one were C/C] (Table 3 of supplementary data). An association *t*-test was used to find genotype-activity correlations but revealed no significant association for any of the four SNPs (data not shown). We concluded that the activity is not modulated by the sequence of the gene, which further supports the findings of our genetic association studies.

DISCUSSION

CD is a complex genetic trait in which genetic and environmental factors are the primary causative determinants for the disease. Although gluten has been identified as the major environmental factor (3), only the genetic contribution of the human leucocyte antigen (HLA) region is well understood (13). Recently, a genome-wide screen in our Dutch population has been successful in finding significant linkage to two non-HLA associated regions, one to chromosome 19 and another to chromosome 6q21–22 (16). No causative gene has been identified yet for either of these regions.

By integrating a data set from our microarray experiments with the genetic information of the 6q21–22 region, we identified eight differentially expressed genes located under this linkage peak. Because one of these differentially expressed genes was *PREP*, we hypothesized that an altered PREP activity in the intestinal mucosa could be responsible for the inefficient breakdown of gluten peptides, which could consequently facilitate the onset of CD. We therefore performed a comprehensive set of complementary studies to investigate the putative role of *PREP* in the pathogenesis of CD.

Because expression studies showed the existence of altered levels of *PREP* mRNA in the biopsies of CD patients, we hypothesized that we might identify a DNA polymorphism or a variant that would slightly alter the activity of the enzyme, rather than a major mutation that would fully abolish its function. Sequence analysis did not reveal any major mutations in 25 CD patients, but six SNPs were found in the coding region of this gene. One of the SNPs is in one of the residues of the catalytic triad (His680), but it does not give rise to an amino acid change. A novel SNP was found in the 5′-UTR of *PREP*. Because the promoter region of *PREP* is not known, in silico studies using Transfac TF professional version 8.2 were used to define whether putative binding sites and regulatory sequences in the 5′-UTR of *PREP* reside at the position of this SNP. No putative regulatory sequence was predicted at the site of the SNP (data not shown), suggesting that this SNP may not affect the transcriptional regulation of *PREP*. SNP rs9486069, located within the first 10 nucleotides of exon 5, was also of potential interest because it has been well established that sequences within the first or last 20 nucleotides of an exon can influence the splicing machinery by enhancing or silencing its effects (1). We therefore looked for a possible influence of this SNP on the splicing machinery using Spring Harbor software (2), but found none (data not shown).

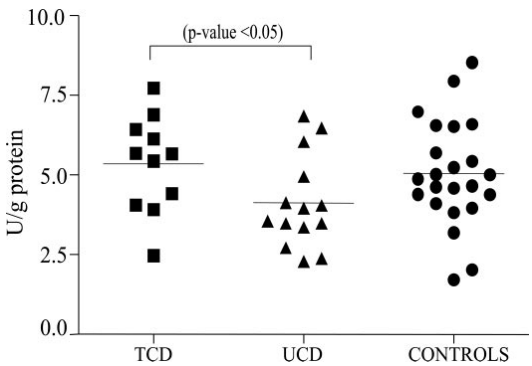


Fig. 2. PREP activity in duodenal biopsies. The activity was measured with Z-Gly-Pro-AMC substrate and corrected for protein concentration determined with BCA assay. TCD 5.35 ± 0.46; UCD 4.12 ± 0.39; $P < 0.05$; controls 5.03 ± 0.033. TCD, treated celiac disease (gluten-free diet); UCD, untreated celiac disease (normal diet); controls, no celiac disease and normal diet.

From the sequence and follow-up analysis we concluded that none of the SNPs would directly provoke a change in the structure of the protein. Neither did our later genetic studies support a role for *PREP* as a primary gene in CD. The microsatellite marker and the six SNPs inside *PREP* did not show any significant differences nor any trend towards significance. Besides, because the promoter region of the *PREP* gene is unknown, SNPs in this region could not be totally excluded.

Finally, to further exclude any functional consequence of these coding polymorphisms in PREP activity that could implicate it in the pathogenesis of CD, we determined the catalytic activity of PREP in biopsies from 47 children. As expected from the genetic association studies, we found no significant differences between the treated or untreated CD children and the pediatric controls or when individuals were grouped by their genotypes. Nevertheless, because the biopsies for normal controls came from individuals who might have had altered intestinal mucosa due to diarrhea or abdominal pain, these results could be an underestimation. However, the PREP activity in untreated CD children was slightly decreased compared with treated CD pediatric patients, possibly as the result of intestinal tissue damage associated with the disease. These observations are perfectly in line with findings of Donlon and Stevens (J. Donlon and F. M. Stevens, personal communication) but do not support results published by Matysiak-Budnik et al. (9), who described an increased PREP activity in the intestinal mucosa of eight treated (i.e., following a gluten-free diet) CD patients compared with seven controls. It remains to be established why our results differ from those of Matysiak-Budnik.

In conclusion, these results clearly indicate that no genetic polymorphisms in the *PREP* gene can be linked to CD. This finding is further supported by the activity determinations in which we found no differences in the enzyme activity between CD patients and controls. Thus *PREP* does not seem to be implicated in the pathogenesis of CD.

ACKNOWLEDGMENTS

The authors thank Ellen van Koppen and Remi Steens for the help in acquiring patient data; Jackie Senior for editing the text; and Alfons Bardoel, Daniel Chan, and Alexandra Zhernakova for the practical work.

GRANTS

This study was supported by the Netherlands Organization for Scientific Research Grants 902-22-094 and 912-02-028, the Celiac Disease Consortium, an innovative cluster approved by The Netherlands Genomics Initiative and partially funded by Dutch Government Grant BSIK03009, and Dutch Digestive Disease Foundation Grant WS00-13.

REFERENCES

1. **Cartegni L, Chew SL, and Krainer AR.** Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3: 285–298, 2002.
2. **Cartegni L, Wang J, Zhu Z, Zhang MQ, and Krainer AR.** ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31: 3568–3571, 2003.
3. **Dicke WKWH, van de Kamer JH.** The presence in wheat of a factor having a deleterious effect in cases of coeliac disease. *Acta Paediatr* 42: 34–42, 1953.
4. **Dieterich W, Esslinger B, and Schuppan D.** Pathomechanisms in celiac disease. *Int Arch Allergy Immunol* 132: 98–108, 2003.
5. **Diosdado B, Wapenaar MC, Franke L, Duran KJ, Goerres MJ, Hadithi M, Crusius JB, Meijer JW, Duggan DJ, Mulder CJ, Holstege FC, and Wijmenga C.** A microarray screen for novel candidate genes in coeliac disease pathogenesis. *Gut* 53: 944–951, 2004.
6. **Franke L, Van Bakel H, Diosdado B, Van Belzen M, Wapenaar M, and Wijmenga C.** TEAM: a tool for the integration of expression, and linkage and association maps. *Eur J Hum Genet*, 2004.
7. **Fulop V, Bocskei Z, and Polgar L.** Prolyl oligopeptidase: an unusual β-propeller domain regulates proteolysis. *Cell* 94: 161–170, 1998.
8. **Goossens F, De Meester I, Vanhoof G, and Scharpe S.** Distribution of prolyl oligopeptidase in human peripheral tissues and body fluids. *Eur J Clin Chem Clin Biochem* 34: 17–22, 1996.
9. **Matysiak-Budnik T, Candalh C, Dugave C, Namane A, Cellier C, Cerf-Bensussan N, and Heyman M.** Alterations of the intestinal transport and processing of gliadin peptides in celiac disease. *Gastroenterology* 125: 696–707, 2003.
11. **Senger S, Luongo D, Maurano F, Mazzeo MF, Siciliano RA, Gianfrani C, David C, Troncone R, Auricchio S, and Rossi M.** Intranasal administration of a recombinant α-gliadin down-regulates the immune response to wheat gliadin in DQ8 transgenic mice. *Immunol Lett* 88: 127–134, 2003.
12. **Shan L, Molberg O, Parrot I, Hausch F, Filiz F, Gray GM, Sollid LM, and Khosla C.** Structural basis for gluten intolerance in celiac sprue. *Science* 297: 2275–2279, 2002.
13. **Sollid LM.** Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* 2: 647–655, 2002.
14. **Vader LW, Stepniak DT, Bunnik EM, Kooy YM, de Haan W, Drijfhout JW, Van Veelen PA, and Koning F.** Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. *Gastroenterology* 125: 1105–1113, 2003.
15. **Vader W, Kooy Y, Van Veelen P, De Ru A, Harris D, Benckhuijsen W, Pena S, Mearin L, Drijfhout JW, and Koning F.** The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* 122: 1729–1737, 2002.
16. **Van Belzen MJ, Meijer JW, Sandkuijl LA, Bardoel AF, Mulder CJ, Pearson PL, Houwen RH, and Wijmenga C.** A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* 125: 1032–1041, 2003.
17. **Vanhoof G, Goossens F, Hendriks L, De Meester I, Hendriks D, Vriend G, Van Broeckhoven C, and Scharpe S.** Cloning and sequence analysis of the gene encoding human lymphocyte prolyl endopeptidase. *Gene* 149: 363–366, 1994.
18. **Venalainen JI, Juvonen RO, and Mannisto PT.** Evolutionary relationships of the prolyl oligopeptidase family enzymes. *Eur J Biochem* 271: 2705–2715, 2004.
19. **Wapenaar MC, van Belzen MJ, Fransen JH, Sarasqueta AF, Houwen RH, Meijer JW, Mulder CJ, and Wijmenga C.** The interferon gamma gene in celiac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. *J Autoimmun* 23: 183–190, 2004.
20. **Working Group of European Society of Paediatric Gastroenterology and Nutrition.** Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. *Arch Dis Child* 65: 909–911, 1990.

Part 3

Increased permeability of the intestinal barrier

Chapter 3.1

Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect

Alienke J Monsuur, Paul I W de Bakker, Behrooz Z Alizadeh, Alexandra Zhernakova, Marianna R Bevova, Eric Strengman, Lude Franke, Ruben van't Slot, Martine J van Belzen, Ineke C M Lavrijsen, Begoña Diosdado, Mark J Daly, Chris J J Mulder, M Luisa Mearin, Jos W R Meijer, Gerrit A Meijer, Erica van Oort, Martin C Wapenaar, Bobby P C Koeleman & Cisca Wijmenga

Nature Genetics 2005;37(12):1341-4

nature genetics

Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect

Alienke J Monsuur[1], Paul I W de Bakker[2], Behrooz Z Alizadeh[1], Alexandra Zhernakova[1], Marianna R Bevova[1], Eric Strengman[1], Lude Franke[1], Ruben van't Slot[1], Martine J van Belzen[1,7], Ineke C M Lavrijsen[1], Begoña Diosdado[1], Mark J Daly[2], Chris J J Mulder[3], M Luisa Mearin[4], Jos W R Meijer[5], Gerrit A Meijer[6], Erica van Oort[1], Martin C Wapenaar[1], Bobby P C Koeleman[1] & Cisca Wijmenga[1]

Celiac disease is probably the best-understood immune-related disorder. The disease presents in the small intestine and results from the interplay between multiple genes and gluten, the triggering environmental factor[1]. Although HLA class II genes explain 40% of the heritable risk, non-HLA genes accounting for most of the familial clustering have not yet been identified. Here we report significant and replicable association ($P = 2.1 \times 10^{-6}$) to a common variant located in intron 28 of the gene myosin IXB (*MYO9B*), which encodes an unconventional myosin molecule that has a role in actin remodeling of epithelial enterocytes[2,3]. Individuals homozygous with respect to the at-risk allele have a 2.3-times higher risk of celiac disease ($P = 1.55 \times 10^{-5}$). This result is suggestive of a primary impairment of the intestinal barrier in the etiology of celiac disease, which may explain why immunogenic gluten peptides are able to pass through the epithelial barrier.

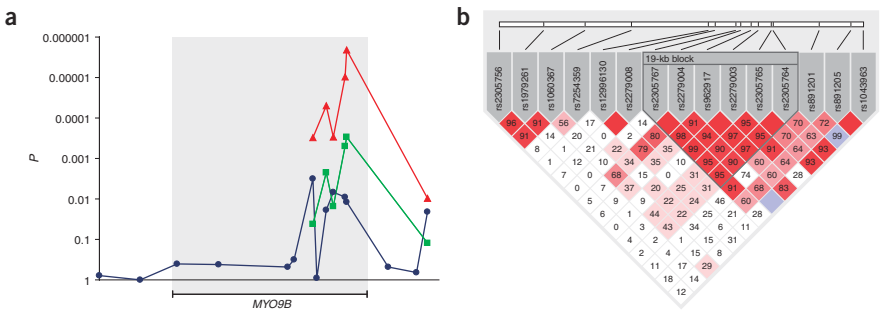


Figure 1 Fine-mapping study in and surrounding *MYO9B*. (**a**) The association data are plotted in blue (circles) on a logarithmic scale for each of the 15 SNPs across a 191-kb region (17,030,936–17,222,116 bp) encompassing *MYO9B* and tested in 216 individuals with Marsh III celiac disease and 216 controls (set 1). Subsequent follow-up studies in a second cohort of 247 affected individuals and 470 controls (set 2) of the six SNPs with $P < 0.05$ are plotted in green (squares). The association data from the combined cohorts (463 affected individuals and 686 controls) are plotted in red (triangles). The spacing between SNPs reflects the distances between them. (**b**) Pairwise LD between the 15 SNPs, given by the D' statistics computed with the genotype data from the 216 controls. The overall LD structure of these data is very similar to that of the data on a European population in the HapMap database (Centre d'Etude du Polymorphisme Humain; Utah residents with ancestry from northern and western Europe), where all associated SNPs fall into a single 22-kb haplotype block (defined as previously described[24]). The darker shade of red indicates the higher D'.

[1]Complex Genetics Section, DBG-Department of Medical Genetics, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, the Netherlands. [2]Broad Institute of Harvard and MIT, 320 Charles St., Cambridge, Massachusetts 02141, USA; and Center for Human Genetics Research, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA. [3]Department of Gastroenterology, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands. [4]Department of Pediatrics, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, and VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands. [5]Department of Pathology, Rijnstate Hospital, P.O. Box 9555, 6800 TA Arnhem, The Netherlands. [6]Department of Pathology, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands. [7]Present address: Laboratory of DNA Diagnostics, Leiden University Medical Center, P.O. Box 9503, 2300 RA Leiden, The Netherlands. Correspondence should be addressed to C.W. (t.n.wijmenga@med.uu.nl).

Received 29 June; accepted 27 September; published online 13 November 2005; doi:10.1038/ng1680

Table 1 Names, chromosomal locations and *P* values for the 15 SNPs tested in and around *MYO9B*

SNP	rs2305756	rs1979261	rs1060367	rs7254359	rs12996130	rs2279008	rs2305767	rs2279004	rs962917	rs2279003	rs2305765	rs2305764	rs891201	rs891205	rs1043963
Position (bp)	17,030,936	17,054,518	17,076,202	17,100,566	17,140,752	17,144,303	17,155,296	17,157,679	17,163,247	17,167,031	17,173,992	17,174,833	17,199,099	17,215,586	17,222,116
P															
Set 1	0.7836	0.9885	0.4085	0.4198	0.4741	0.3114	0.0032*	0.9049	0.0186	0.0069	0.0089	0.0121	0.4832	0.6614	0.0205
Set 2							0.0412		0.0022*	0.0151	0.0005*	0.0003*			0.1233
Set 1 and set 2							0.0003*		0.000049*	0.0003*	0.0000097*	0.0000021*			0.0095

* *P* values still significant after correcting for multiple testing. *P* values were calculated in Haploview using the χ^2 test. The locations of the SNPs in *MYO9B* are given in **Supplementary Figure 2**.

With a prevalence close to 1% (ref. 4), celiac disease is the most common food intolerance in general western populations. In individuals with celiac disease, ingestion of gluten leads to inflammation and tissue remodeling of the intestinal mucosa, resulting in malnutrition and severe complications. Long regarded as a gastrointestinal disorder of childhood, the disease is now considered to be a chronic systemic autoimmune disease[5] and is more often diagnosed in adults than in children[6]. Celiac disease has a strong heritable component, although the inheritance is complex and multifactorial[1]. Gluten is the key environmental risk factor, and HLA class II genes are associated to celiac disease[7]. HLA-DQ2 has a key role in the disease by presenting gluten peptides to $CD4^+$ cells in the lamina propria[8]. Much of the research on celiac disease has focused on the activation and regulation of gluten-reactive T cells. But the etiological steps preceding this T-cell activation are still poorly understood; for example, why are the antigenic gluten peptides resistant to further breakdown in the intestinal lumen, and how do the gluten peptides pass through the epithelial barrier? The integrity of the intestinal barrier is impaired in active celiac disease[9,10], which implies that the epithelial cell barrier has a role in the early pathogenesis of the disease.

Genome-wide screens have been done to identify non-HLA genes involved in celiac disease but, so far, no gene has been positionally cloned. We previously obtained strong evidence for linkage (multiple maximum lod score 4.43, nominal $P = 6.2 \times 10^{-6}$) to chromosome 19p13.1 in celiac disease (*CELIAC4*) in a cohort of affected sibling pairs of European descent from the Netherlands[11]. This chromosomal location was also suggested by meta and pooled analyses of European celiac disease data that did not include the Dutch cohort[12]. In our original study[11], further association analysis was done with five microsatellite markers spanning the 3.5-Mb lod-1 interval in a cohort of 216 case-control pairs, resulting in weak association to *D19S899* ($P = 1.3 \times 10^{-3}$), located in intron 1 of the gene myosin IXB (*MYO9B*).

Here we focused on the region surrounding *D19S899* and, more specifically, on *MYO9B* using the same 216 cases and 216 controls (set 1; **Supplementary Fig. 1** and **Supplementary Table 1** online). Initially, we typed 15 SNPs with an average spacing of 13 kb across *MYO9B*. Using a conservative Bonferroni correction, a single SNP (rs2305767) located in intron 14 of *MYO9B* showed significant association ($P = 0.0032$; **Fig. 1a**, **Table 1** and **Supplementary Table 2** online), confirming our initial observation. Another five SNPs proximal to rs2305767 also showed significant association without correction ($P < 0.05$; **Table 1**); these SNPs allowed us to refine the region of association, as they were located in the 3′ part of *MYO9B*, with one SNP (rs1043963) located 36.4 kb downstream of the gene. A replication study of the six SNPs with $P < 0.05$ was done in a second, fully independent, cohort of 247 unrelated individuals with Marsh III celiac disease who met the same strict diagnostic criteria[11], and 470 Dutch blood bank donors (set 2). We observed significant association ($P < 0.05$) for all five SNPs in *MYO9B* but not for the SNP downstream of the gene (rs1043963, $P = 0.123$). Three SNPs (rs962917, rs2305765 and rs2305764) were still significant after correcting for six independent tests (nominal Bonferroni $P < 0.05/6 = 8 \times 10^{-3}$; **Fig. 1** and **Table 1**). Combining the cohorts (set 1 and set 2) strengthened the association considerably and showed highly significant association for all five SNPs in *MYO9B* (**Fig. 1** and **Table 1**). We observed the smallest *P* value for rs2305764 ($P = 2.1 \times 10^{-6}$) located in intron 28. Our data showed strong linkage disequilibrium (LD) between the five associated SNPs (**Fig. 1b**), which are located in a single, 19-kb haplotype block spanning exons 15–27 of *MYO9B* (**Supplementary Fig. 2** online).

To rule out the possibility that our initial study (comprising only five microsatellite markers[11]) had overlooked an association signal elsewhere in the *CELIAC4* locus, we carried out a comprehensive fine-mapping study using 359 tag SNPs with minor allele frequencies > 0.02 and $r^2 > 0.7$, and covering the lod-1.5 region (99% confidence interval (c.i.); 15,385,880–21,075,237 bp) on set 2 supplemented with the cases from set 1 (totaling 463 independent cases and 470 controls). After carrying out quality checks on the raw data, we excluded 60 poorly performing or monomorphic tag SNPs (**Supplementary Table 3** online), yielding 299 tag SNPs for further analysis, of which 19 were located in the 191-kb region covered previously. Three SNPs were overlapping in the two screens (rs7254359, rs2279008 and rs2305767; **Supplementary Table 2**). After excluding tag SNPs that showed deviation from Hardy-Weinberg equilibrium, we were able to analyze 291 tag SNPs successfully. These allowed us to capture most of the untyped genetic variation present in the Centre d'Etude du Polymorphisme Humain population of the HapMap Project (**Supplementary Table 3**). We observed a single, strong peak of association in *MYO9B* (**Fig. 2**, **Table 2** and **Supplementary Table 3**). The most-associated tag SNP (rs1457092, $P = 7.8 \times 10^{-5}$), located between rs962917 and rs2279003, remained significant after stringent Bonferroni correction for 291 tag SNPs ($P_c < 0.05/291 = 1.72 \times 10^{-4}$) and is in strong LD with the five associated SNPs identified in the random screen described above (rs2305767, rs962917, rs2279003, rs2305765

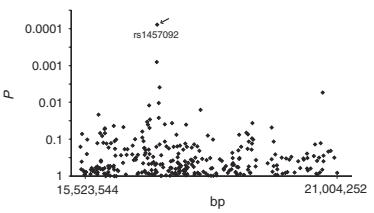


Figure 2 Comprehensive fine-mapping study of the *CELIAC4* region on 19p13.1. *P* values are plotted for all 291 tag SNPs within the 99% c.i. (15,385,880–21,075,237 bp, in chromosomal order) in a cohort of 463 independent individuals with Marsh III celiac disease and 470 controls. *P* values were calculated in Haploview using the χ^2 test. The *P* value cutoff after Bonferroni correction was 1.72×10^{-4}. Only SNP rs1457092, located in intron 20 of *MYO9B* and indicated with an arrow, was significant after correction for multiple testing.

Table 2 Names, chromosomal locations and *P* values for the 11 tag SNPs from the comprehensive screen located in *MYO9B* (17,072,928–17,185,703 bp)

SNP	rs7246865	rs4808571	rs7254359	rs1870068	rs10409461	rs11673417	rs2279009	rs2279008	rs7259292	rs2305767	rs1457092
Position (bp)	17,080,105	17,087,229	17,100,566	17,119,828	17,129,784	17,131,499	17,143,899	17,144,303	17,153,014	17,155,296	17,165,236
P	0.3022	0.4125	0.9009	0.4977	0.7905	0.6813	0.4592	0.1617	0.7726	0.0008	0.000078*

P* value was still significant after correcting for multiple testing. Details are shown in **Supplementary Table 3.

and rs2305764; **Supplementary Fig. 3** online). One of the three overlapping SNPs (rs2305767 located in intron 14) was already significant in the initial cohort ($P = 3.2 \times 10^{-3}$) and showed even stronger association in the combined cohorts (set 1 and set 2, comprising 463 cases and 686 controls; $P = 3 \times 10^{-4}$; **Table 1**).

Haplotype analysis with these six SNPs showed that in this 19-kb region, only four haplotypes account for >97% of all the observed haplotypes; these can be captured by typing only three SNPs (rs2305767, rs1457092 and rs2305764). The AAACTA haplotype was present in 39% of cases and 30% of controls (odds ratio (OR) = 1.56; 95% c.i. = 1.27–1.93, $P = 1.8 \times 10^{-5}$; **Table 3**). The GGCTCG haplotype is the most low-risk haplotype and is very unlikely to encompass the unknown causal disease variant. We were able to narrow the haplotype to a single SNP, as we observed that the presence of the A allele of SNP rs2305764 constantly predisposes to a higher risk of celiac disease, independent of the variation in the other five SNPs. This observation was confirmed by logistic regression analysis using a conditional forward selection procedure showing that only SNP rs2305764 was left in the regression model (data not shown) and, similarly, that none of the haplotypes showed stronger association. Therefore, the rs2305764 SNP located in intron 28 of *MYO9B* can alone completely explain the association observed at *MYO9B* and can be considered a marker for celiac disease risk. Sequence analysis of the entire *MYO9B* coding region in 16 affected individuals homozygous with respect to the at-risk haplotype identified 22 variants, 2 of which were already included in our study. Of the 20 new SNP variants, 2 were considered relevant (*i.e.*, nonsynonymous SNPs) and were typed in our entire cohort of 463 independent cases and 686 independent controls. Unfortunately, neither of these two SNPs turned out to be more strongly associated than rs2305764, which agrees with the observation that these two SNPs are not in LD with the at-risk haplotype (**Supplementary Table 4** and **Supplementary Fig. 4** online).

The genotype frequency data for rs2305764 showed that the risk of celiac disease rose in proportion to the number of copies of the A allele (frequency of 46.5% in affected individuals compared with 37.9% in controls), implying a codominant allele-dose effect. Individuals heterozygous with respect to the A allele have a modest but significantly higher risk of developing celiac disease (OR = 1.66; 95% c.i. = 1.23–2.13; $P = 5.3 \times 10^{-4}$), whereas individuals homozygous with respect to the A allele have a risk of developing celiac disease that increases to 2.27 (95% c.i. = 1.56–3.30; $P = 1.55 \times 10^{-5}$), with population-attributable risks of 25% and 23%, respectively. These results suggest that the 3′ part of *MYO9B* is associated with celiac disease.

MYO9B is a good candidate gene for celiac disease because of the function of its encoded protein, and it may be one of the long-sought factors involved in the early mucosal events preceding the well-understood inflammatory response. *MYO9B* encodes a single motor protein[2] belonging to the class IX myosin molecules, which are unique in comparison with other classes because they contain a Rho-GTPase-activating domain within their tails. This GTPase activity converts active Rho-GTP into inactive Rho-GDP, thereby downregulating Rho-dependent signaling pathways[3]. Rho-family GTPases are involved in remodeling of the cytoskeleton and tight junction assembly, both of which result in enhanced epithelial paracellular permeability[13,14]. It is therefore tempting to speculate that a genetic variant in the 3′ part of *MYO9B* leads to an impaired interaction with RhoA, thereby perturbing tight junction gate and fence function. Hence, a subtle, underlying intestinal barrier abnormality may be involved in the etiology of celiac disease, which is in line with the recent observation of intestinal permeability in individuals with celiac disease with normal histology[15,16]. As a consequence, immunogenic gluten peptides can enter the deeper mucosal layer more easily. Notably, this is the site at which the HLA-DQ2–mediated antigen presentation to the $CD4^+$ cells initiates the inflammatory response. So far, *MYO9B* is the only non-HLA gene identified for celiac disease by positional cloning. It will be interesting to determine the effect of this gene in other populations with celiac disease. The identification of *MYO9B* as a susceptibility gene in celiac disease is a notable finding that may open new avenues for studying the early events of celiac disease pathogenesis, a process that has not yet received much attention, but which might prove important in developing alternative treatments to the strict gluten-free diet currently used.

METHODS

Subjects and controls. DNA, isolated from whole blood, was available from two independent cohorts of Dutch individuals with celiac disease (set 1, 216; set

Table 3 The prevalence of *MYO9B* haplotypes reconstructed from selected SNPs and their association to celiac disease

MYO9B haplotypes										
rs2305767	rs962917	rs1457092	rs2279003	rs2305765	rs2305764	Cases (%)[a]	Controls (%)[a]	OR[b]	95% c.i.[b]	*P*[c]
G	G	C	T	C	G	308 (37.3)	565 (45.5)	1.00[d]	–	–
A	G	C	C	C	G	128 (15.5)	216 (17.4)	1.09	0.84–1.41	0.53
A	A	A	C	T	A	319 (38.7)	375 (30.2)	1.56	1.27–1.93	0.000018
A	G	C	C	T	A	46 (5.6)	67 (5.4)	1.26	0.84–1.88	0.26
Rare haplotypes						24 (2.9)	20 (1.6)	2.21	1.24–4.05	0.009

[a]Number represents the frequency of haplotypes estimated using an expectation maximization algorithm embedded log linear model[21,22]. [b]OR represents maximum likelihood estimate of odds ratio, and the corresponding 95% c.i. was approximated using Woolf's method. [c]*P* values were calculated using a χ^2 test. [d]This haplotype was taken as the reference.

2, 247). All the affected individuals were diagnosed in accordance with the revised ESPGHAN criteria[17]. More than 90% of the affected individuals were HLA-DQ2-positive (**Supplementary Table 1**). The initial biopsy specimens of the individuals were retrieved; all showed a Marsh III lesion upon reevaluation by one of two experienced pathologists (G.A.M. and J.W.R.M.). Both cohorts included children and adults. There were also two cohorts of controls available. Set 1 controls ($n = 216$) comprised random hospital controls. Set 2 controls ($n = 470$) were random blood bank donors. All cases and controls were from The Netherlands and of European descent, and at least three of their four grandparents were also born in The Netherlands. The comprehensive screen included all individuals in set 2 supplemented with the cases from set 1. **Supplementary Table 1** lists the baseline demographic parameters of the cohorts. All individuals gave their informed consent. This study was approved by the Medical Ethical Committee of the University Medical Center Utrecht.

Random screen: SNP selection and genotyping. We selected 15 random SNPs from Applied Biosystems, covering *MYO9B* and its surroundings. These SNPs were obtained as Assays on Demand or Assays by Design (Applied Biosystems) and initially typed on set 1 (216 individuals with Marsh III celiac disease and 216 controls) using a 7900 Taqman (Applied Biosystems). In the follow-up study, six SNPs were selected for further typing on set 2 (247 individuals with Marsh III celiac disease and 470 controls).

Comprehensive screen: tag SNP selection and genotyping. For the comprehensive screen we took the 99% c.i. (1.5-lod) of the *CELIAC4* linkage peak on chromosome 19, ranging from 15,385,880 to 21,075,237 bp (National Center for Biotechnology Information build 34). SNPs were selected by downloading all the SNPs typed in the Centre d'Etude du Polymorphisme Humain (Utah residents with ancestry from northern and western Europe) population in this region from the HapMap database[18]. From these SNPs, the program Tagger[19] was used to select tag SNPs so that all SNPs with a minor allele frequency $\geq 2\%$ were captured with $r^2 \geq 0.7$ (excluding SNPs with low Illumina quality design scores). A final set of 359 tag SNPs was obtained for genotype analysis.

We genotyped SNPs using the GoldenGate assay on an Illumina BeadStation 500 GX (Illumina). All tag SNPs were examined for their resulting quality; all those that were not polymorphic in our population, had a low signal or had clusters that were too wide were excluded (**Supplementary Table 3**). We also carried out an evaluation of the tagging efficiency for two reasons. First, several tag SNPs were dropped because of low-quality or bad clusters, because they were not in Hardy-Weinberg equilibrium in the controls, or because they were not polymorphic in our population. Second, in the meantime, HapMap had expanded the number of SNPs typed in the region. We evaluated the tagging efficiency of the 291 working tag SNPs using the following parameters (**Supplementary Table 3**): aggressive tagging was used to get the most information; all the SNPs in the region were used; lod score threshold was set to 1; number of iterations was set to 1,000; maximum number of tries was set to 100,000.

Statistical analysis. We calculated association χ^2 and two-tailed P values using the Haploview program[20], for each stage of the study (set 1, set 2, set 1 and set 2, and the comprehensive screen). SNPs that were not in Hardy-Weinberg equilibrium ($P < 0.001$) in the controls were excluded from further analysis. We used multiple logistic regression analysis to estimate allelic and genotypic OR and the corresponding 95% c.i. for SNP rs2306764. To obtain genotype information for the combined cohorts on the six SNPs comprising this 19-kb region and showing $P < 10^{-3}$ (rs2305767, rs962917, rs1457092, rs2279003, rs2305765 and rs2305764), the associated tag SNP (rs1457092) was also typed for the 216 controls from set 1. As parental information was missing, an expectation maximization algorithm was used to estimate haplotype frequencies in a multiple locus system[21,22]. This function recast the regression model as a generalized linear mixed model with random effects, which is fitted as part of the maximization step and allows testing for LD and disease association[21,22]. The OR and the corresponding 95% c.i. were calculated for all haplotypes. Population attributable risk was calculated for rs2306764 using the corresponding genotype frequencies of this marker as explained previously[23]. All the analyses were done using STATA statistical software, version 8.0 for MS Windows.

URLs. The website for Applied Biosystems is http://myscience.appliedbiosystems.com/. The program Tagger is available at http://www.broad.mit.edu/mpg/tagger/. The HapMap database and the Haploview program are available at http://www.hapmap.org/.

Accession codes. GenBank: *Homo sapiens* chromosome 19 complete sequence, NC_000019; *H. sapiens MYO9B* mRNA, NM_004145. Ensembl: transcript *MYO9B*, ENST00000319396.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the affected individuals, their physicians and the Dutch Celiac Disease Foundation for participating in this study; C. Gouw, A. Bardoel and H. van Someren for practical assistance and J. Senior for critically reading the manuscript. This study was supported by grants from The Netherlands Organization for Scientific Research, the Dutch Digestive Disease Foundation and the Celiac Disease Consortium.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Van Heel, D.A., Hunt, K., Greco, L. & Wijmenga, C. Genetics in coeliac disease. *Best Pract. Res. Clin. Gastroenterol.* **19**, 323–339 (2005).
2. Post, P.L. *et al.* Myosin-IXb is a single-headed and processive motor. *J. Biol. Chem.* **277**, 11679–11683 (2002).
3. Muller, R.T., 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* **8**, 2039–2053 (1997).
4. Dube, C. *et al.* The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. *Gastroenterology* **128**, S57–S67 (2005).
5. Rewers, M. Epidemiology of celiac disease: what are the prevalence, incidence, and progression of celiac disease? *Gastroenterology* **128**, S47–S51 (2005).
6. Green, P.H. The many faces of celiac disease: clinical presentation of celiac disease in the adult population. *Gastroenterology* **128**, S74–S78 (2005).
7. Falchuk, Z.M. & Strober, W. HL-A antigens and adult coeliac disease. *Lancet* **2**, 1310 (1972).
8. Arentz-Hansen, H. *et al.* The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. *J. Exp. Med.* **191**, 603–612 (2000).
9. Schulzke, J.D., Schulzke, I., Fromm, M. & Riecken, E.O. Epithelial barrier and ion transport in coeliac sprue: electrical measurements on intestinal aspiration biopsy specimens. *Gut* **37**, 777–782 (1995).
10. Van Elburg, R.M., Uil, J.J., Mulder, C.J. & Heymans, H.S. Intestinal permeability in patients with coeliac disease and relatives of patients with coeliac disease. *Gut* **34**, 354–357 (1993).
11. Van Belzen, M.J. *et al.* A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* **125**, 1032–1041 (2003).
12. Babron, M.C. *et al.* Meta and pooled analysis of European coeliac disease data. *Eur. J. Hum. Genet.* **11**, 828–834 (2003).
13. Bruewer, M., Hopkins, A.M., Hobert, M.E., Nusrat, A. & Madara, J.L. RhoA, Rac1, and Cdc42 exert distinct effects on epithelial barrier via selective structural and biochemical modulation of junctional proteins and F-actin. *Am. J. Physiol. Cell Physiol.* **287**, C327–C335 (2004).
14. Matter, K. & Balda, M.S. Signalling to and from tight junctions. *Nat. Rev. Mol. Cell Biol.* **4**, 225–236 (2003).
15. Smecuol, E. *et al.* Permeability, zonulin production, and enteropathy in dermatitis herpetiformis. *Clin. Gastroenterol. Hepatol.* **3**, 335–341 (2005).
16. Uil, J.J. *et al.* Follow-up of treated coeliac patients: sugar absorption test and intestinal biopsies compared. *Eur. J. Gastroenterol. Hepatol.* **8**, 219–223 (1996).
17. United European Gastroenterology. When is a coeliac a coeliac? Report of a working group of the United European Gastroenterology Week in Amsterdam, 2001. *Eur. J. Gastroenterol. Hepatol.* **13**, 1123–1128 (2001).
18. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
19. De Bakker, P.I. *et al.* Efficiency and power in genetic association studies. *Nat. Genet.*, advance online publication 23 October 2005 (doi:10.1038/ng1669).
20. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
21. Chiano, M.N. & Clayton, D.G. Fine genetic mapping using haplotype analysis and the missing data problem. *Ann. Hum. Genet.* **62**, 55–60 (1998).
22. Bitti, P.P. *et al.* Association between the ancestral haplotype HLA A30B18DR3 and multiple sclerosis in central Sardinia. *Genet. Epidemiol.* **20**, 271–283 (2001).
23. Greenland, S. & Rothman, R.K. Measures of effect and measures of association. in *Modern Epidemiology* 2nd edn. 47–67 (Lippincott-Raven, Philadelphia, 1998).
24. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).

Chapter 3.2

Lack of association of *MYO9B* genetic variants with coeliac disease in a British cohort

Karen A Hunt, Alienke J Monsuur, WL McArdle, Parveen J Kumar, Simon PL Travis, Julian RF Walters, Derek P Jewell, David P Strachan, Raymond J Playford, Cisca Wijmenga, David A van Heel

Gut 2006; 55(7):969-72

INFLAMMATORY BOWEL DISEASE

Lack of association of MYO9B genetic variants with coeliac disease in a British cohort

K A Hunt, A J Monsuur, W L McArdle, P J Kumar, S P L Travis, J R F Walters, D P Jewell, D P Strachan, R J Playford, C Wijmenga, D A van Heel

Gut 2006;**55**:969–972. doi: 10.1136/gut.2005.086769

See end of article for authors' affiliations

Correspondence to: Professor D A van Heel, Institute of Cell and Molecular Science, Barts and The London, Queen Mary's School of Medicine and Dentistry, Turner St, London E1 2AD; d.vanheel@qmul.ac.uk

Revised version received 19 December 2005
Accepted for publication 22 December 2005
Published online first 19 January 2006

Background and aims: Development of coeliac disease involves an interaction between environmental factors (especially dietary wheat, rye, and barley antigens) and genetic factors (there is strong inherited disease susceptibility). The known human leucocyte antigen (*HLA*)-*DQ2* and -*DQ8* association explains only a minority of disease heritability. A recent study in the Dutch population suggested that genetic variation in the 3′ region of myosin IXB (*MYO9B*) predisposes to coeliac disease. MYO9B is a Rho family GTPase activating protein involved in epithelial cell cytoskeletal organisation. MYO9B is hypothesised to influence intestinal permeability and hence intestinal antigen presentation.
Methods: Four single nucleotide polymorphisms were chosen to tag all common haplotypes of the *MYO9B* 3′ haplotype block (exons 15–27). We genotyped 375 coeliac disease cases and 1366 controls (371 healthy and 995 population based). All individuals were of White UK Caucasian ethnicity.
Results: UK healthy control and population control allele frequencies were similar for all *MYO9B* variants. Case control analysis showed no significant association of any variant or haplotype with coeliac disease.
Conclusions: Genetic variation in *MYO9B* does not have a major effect on coeliac disease susceptibility in the UK population. Differences between populations, a weaker effect size than originally described, or possibly a type I error in the Dutch study might explain these findings.

Coeliac disease, characterised by small intestinal inflammation induced by dietary wheat, rye, and barley, is common in the UK population, with a prevalence of ~1%. Twin and family studies indicate a high inherited predisposition.[1] The only clearly identified genetic risk factor for disease is the human leucocyte antigen (*HLA*)-*DQ* locus. HLA-DQ is critical to present toxic cereal derived peptides to intestinal T cells. However, although the *HLA-DQ2* variant is present in >90% of coeliacs (or DQ8 in a minority), it is also found commonly in healthy individuals (~30%). This suggests that *HLA-DQ2* is necessary but not sufficient for disease development. The *HLA-DQ* locus accounts for <40% of coeliac disease heritability.[1–3]

A recent study suggested that genetic variation in the *MYO9B* gene (encoding myosin IXB) might predispose to coeliac disease.[4] This gene lies within a region of human chromosome 19 first identified in a genome wide linkage study performed in a Dutch population.[5] A meta-analysis of other European linkage studies provided further (albeit weak) support for this finding.[6] Monsuur *et al* recently studied 291 haplotype tagging single nucleotide polymorphisms (SNPs) covering the 99% confidence interval of this linkage peak, and reported significant association in two independent Dutch case control cohorts with variants in the 3′ region of the *MYO9B* gene (which are part of a large haplotype block between exons 15–27).[4] A non-coding variant (A allele of rs2305764) in intron 28 completely explained the observed association and was considered a marker for disease risk. This allele was reported to confer odds ratios for developing coeliac disease of 1.7 (heterozygotes) and 2.3 (homozygotes).

It is currently unknown how this non-coding variant might influence *MYO9B* expression or function. MYO9B is a single motor protein with a Rho GTPase activating domain.[7 8] Rho-GTPase proteins are involved in epithelial cell tight junction assembly and cytoskeletal remodelling.[9] It has been hypothesised that *MYO9B* variants might alter epithelial permeability, thus allowing toxic wheat, rye, and barley components to be presented to the immune system in coeliac disease patients. We sought to assess the impact of *MYO9B* genetic variation in a large cohort of UK coeliac disease patients.

METHODS

Study cohorts

White Caucasian individuals recruited from the UK were studied. Coeliac disease patients (three sites in South East England) were included. Inclusion criteria were as described previously,[10] based on the presence of villous atrophy at diagnosis and (since test introduction) positive antiendomysial/tissue transglutaminase antibody (table 1). Healthy controls were recruited from clinical and laboratory staff volunteers, and from UK National Blood Transfusion Service donors. Population based controls were analysed from the 1958 British Birth Cohort. Ethics committee and local approval were obtained for all cohorts. Genomic DNA was extracted from peripheral blood or from immortalised peripheral blood lymphocyte cell lines (1958 British Birth Cohort).

Genotyping

Three variants (rs2305767, rs2305765, rs2305764) were genotyped using the Sequenom platform (Sequenom, San Diego California, USA). Primers for the multiplex assay were designed using the Sequenom SpectroDesigner software. The Taqman platform (Applied Biosystems, Warrington, UK) was used for rs1457092 genotyping. We also genotyped 342 DNA samples for rs2305767, rs2305765, and rs2305764 with both Taqman and Sequenom methods and obtained identical results. All primer sequences and conditions are available on request. Genotyping on both platforms had >96% success rate. The 384 well plates contained multiple negative controls

Abbreviations: HLA, human leucocyte antigen; SNP, single nucleotide polymorphism; MYO9B, gene encoding myosin IXB

Table 1 Demographics of the UK cohorts and comparison with previously reported Dutch cohorts

	UK coeliac cases	Dutch coeliac cases*	UK healthy controls	UK population controls	Dutch controls†
No of individuals	375	463	371	995	686
Female (%)	73%	69%	48%‡	50%	49%
Age (y) (at Dec 2005) (median (range))	55 (17–86)	44 (3–93)	na (all >16)	48 (48–48)	48 (6–93)
Age at diagnosis (y) (median (range))	42 (0–78)	35 (1–83)			
HLA-DQ2 positive (%) (possession of DQA1*05 + DQB1*02 heterodimer)	94%§	94%	na	na	na
Small intestinal villous atrophy (%) (at diagnosis, Marsh III lesion)	100%	100%			
IgA antiendomysial/tTG antibody status (%)	¶				
Known positive (when untreated)	70.6%	na			
Not performed when untreated	28.2%				
Known IgA deficiency	1.2%				

*Combined cases set 1 and 2 from Dutch study.[4]
†Combined controls set 1 and 2 from Dutch study.[4]
‡Data available on 134 of 371 UK healthy controls, others mostly anonymous blood donors.
§Data available on 249 of 375 UK cases.
¶Serology data available for 344 of 375 UK coeliac cases.
na, not available; tTG, tissue transglutaminase; HLA, human leucocyte antigen.

(no DNA). The 1958 British Birth Cohort samples contained 12 blinded duplicates across genotyping plates. No genotype discrepancies were observed.

A previous study demonstrated that three of these variants (rs2305767, rs2305764, rs1457092) were sufficient to tag 97% of all observed Caucasian haplotypes for the 19 kb block, spanning exons 15–27 of *MYO9B* (this block contains all of the variants showing association in the Dutch population[4]). We genotyped rs2305765 to provide an independent test of the most associated SNP in the Dutch population (rs2305764). We confirmed these two markers to be in almost complete linkage disequilibrium in the UK population, as reported in the Dutch population.[4]

Statistics

Case control association analyses were performed using the Haploview program (version 3.2).[11] Haplotypes were estimated from the unphased individual data using an accelerated expectation-maximisation algorithm.[12] Association was tested using frequency counts in a two tailed χ^2 test for both individual SNPs and common (>1%) haplotypes. Significance was assumed at $p<0.05$. All p values are presented uncorrected for multiple statistical testing. Power calculations were performed using the genetic power calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/).[13] Genotype data for the four variants in both control groups was tested (Haploview) and found to be in Hardy-Weinberg equilibrium ($p>0.01$).

RESULTS

We first estimated that our study had 94% power at $p<0.001$ to detect association in the UK coeliac disease and population control cohorts (assuming similar results as reported in the Dutch population for rs2305764 with an allele frequency of 0.379, odds ratio 1.66 for AG genotype, and odds ratio 2.27 for AA genotype). In a ''worst case'' scenario, assuming the lower 95% confidence intervals of genotypic odds ratios reported in the Dutch population, our study had 72% power at $p<0.05$ to detect association.

Table 2 summarises the genotyping results and case control association analyses at the single SNP level. Allele frequencies for all four *MYO9B* variants studied were similar in the UK healthy control and population control populations, as expected. No significant difference in allele frequencies was observed between males (n = 500) and females (n = 495) from the population control cohort (data not shown). No significant association (uncorrected $p<0.05$) was seen with coeliac disease cases compared with healthy or population controls. No significant association was seen between coeliac disease cases and combined control cohorts.

Modest differences in allele frequencies of rs2305764 (reported to be the strongest associated *MYO9B* variant in Dutch coeliac disease[4]) were observed between populations: Dutch healthy controls (37.9%), UK healthy controls (42.3%), and UK population controls (41.1%); Dutch coeliac disease (46.5%) and UK coeliac disease (41.8%).

Table 3 summarises the results of haplotype estimation and haplotype based case control association analyses. Haplotype

Table 2 Case control association analysis of *MYO9B* (gene encoding myosin IXB) variants by allele

Cohort	SNP	Minor/major allele	n with genotypes	Homozygous (minor allele)	Heterozygous	Homozygous (major allele)	Minor allele frequency (%)	p Value v healthy controls	p Value v population controls
Coeliac disease	rs2305767	G/A	348	65	161	122	41.8	0.61	0.71
(n = 375)	rs1457092	A/C	356	41	170	145	35.4	0.28	0.31
	rs2305765	T/C	371	55	178	138	38.8	0.25	0.22
	rs2305764	A/G	373	55	179	139	38.7	0.16	0.26
Healthy controls	rs2305767	G/A	352	51	183	118	40.5	–	0.82
(n = 371)	rs1457092	A/C	347	46	173	128	38.2	–	0.76
	rs2305765	T/C	364	58	188	118	41.8	–	0.88
	rs2305764	A/G	366	60	190	116	42.3	–	0.57
Population controls	rs2305767	G/A	971	181	434	356	41.0	–	–
(n = 995)	rs1457092	A/C	971	146	437	388	37.5	–	–
	rs2305765	T/C	939	175	428	336	41.4	–	–
	rs2305764	A/G	977	177	450	350	41.1	–	–

All p values are presented without correction for multiple testing.
SNP, single nucleotide polymorphism.

Table 3 Case control association analysis of *MYO9B* (gene encoding myosin IXB) variants by haplotype in UK cohorts

MYO9B SNP haplotype				Cohort*		
rs2305767	rs1457092	rs2305765	rs2305764	Coeliac disease	Healthy controls	Population controls
G	C	C	G	41.7%	39.0% p=0.51	40.3% p=0.31
A	A	T	A	34.7%	37.6% p=0.34	36.7% p=0.21
A	C	C	G	19.3%	17.9% p=0.42	18.0% p=0.52
A	C	T	A	3.8%	3.5% p=0.96	3.8% p=0.81

Rare (<1%) haplotypes are not shown.
All p values are presented without correction for multiple testing.

frequencies for all four *MYO9B* variants studied were similar in the UK healthy control and population control populations. No significant association (uncorrected $p<0.05$, individual values not shown) was seen for coeliac disease cases compared with healthy or population controls.

DISCUSSION

Our understanding of the pathogenesis of coeliac disease is now better than for most other human autoimmune disorders. Recent advances include identification of immunologically dominant wheat T cell epitopes, the role of tissue transglutaminase, and how HLA-DQ2 binds wheat peptides.[14] However, the primary genetic causes outside of the HLA remain largely unknown. Several factors suggest that identification of disease predisposing susceptibility genes should be possible: the disease shows strong heritability, the disease is common (1% prevalence in UK population), and there are highly sensitive and specific diagnostic tests. The HLA locus was recognised as being involved in coeliac disease pathogenesis in the early 1970s but identification of other disease susceptibility genes has been elusive.

A recent study in the Dutch population[4] identified genetic variation in the 3′ region of *MYO9B* as a risk factor for coeliac disease. This study reported significant association with multiple variants in the 3′ region of the *MYO9B* gene, specifically found within a haplotype block between exons 15–27.[4] The association observed in the *MYO9B* gene could be pinpointed to a single SNP (rs2305764, a non-coding variant in intron 28), which could be considered a marker for coeliac disease risk. Although part of the function of MYO9B is understood, it is not clear how this genetic variant might alter MYO9B activity. Indeed, it is possible that the actual disease causing mutation lies elsewhere within *MYO9B* or other close at hand genes, and has not been located by initial resequencing efforts.

We have analysed a large cohort of coeliac patient samples and two sets of control samples to assess the role of *MYO9B* variants in the British population. Our study found no evidence to support the recent findings in the Dutch population.[4] Indeed, the Dutch study found a higher frequency of the rs2305764 A allele in coeliac disease compared with controls whereas the frequency of the A allele was (non-significantly) lower in UK coeliac disease cases compared with controls. We did not comprehensively assess common genetic variation in the entire *MYO9B* gene, which spans 111 575 base pairs and contains five haplotype blocks.[4]

There are a number of potential reasons for the discrepancy between the UK and Dutch studies which are worthy of specific discussion in context here (although most are well recognised in complex genetic trait studies[15]).

- *Explanation 1: The Dutch report is a false positive*. The study by Monsuur and colleagues[4] involved multiple association testing of a large number of several hundred genetic markers, with a second independent case control cohort used to replicate initial findings and so minimise type I error. The association signal (greatest at $p = 2.1\times10^{-6}$ for rs2305764) withstood conservative Bonferroni correction. Nevertheless, despite these strategies, only reports testing *MYO9B* variants in multiple independent populations by other independent investigators will resolve the possibility of type I error.
- *Explanation 2: The current UK report is a false negative*. We examined the statistical power of the current study using genotypic odds ratio results, as published in the Dutch population, and concluded that our study was highly powered. Our study had similar numbers of coeliac cases (375 versus 463 in the Dutch study) and twice as many controls (1346 versus 686 in the Dutch study). If, however, the actual effect size of *MYO9B* was at the lower confidence limits of that reported in the Dutch population, our study was considerably weaker although power was just within accepted bounds. The effect size observed in a initial genetic study is commonly stronger than subsequent reports, as illustrated by the studies of the *NOD2* and *IBD5/5q31* variants in Crohn's disease.[16–19]
- *Explanation 3: There is heterogeneity between British and Dutch populations*. It is possible, although highly unlikely, that the reported coeliac disease association in the Dutch study is due to a founder effect in the Dutch population. Coeliac disease is reported to have similar clinical features, immunopathogenesis (for example, toxic T cell epitopes, serological responses) and prevalence in the UK and Dutch populations. Furthermore, similar frequencies of *NOD2* mutations have been found in Crohn's disease patients from Caucasian populations in the UK, across Europe, the USA, and Australia. We specifically compared demographic and diagnostic data between the British and previously reported Dutch cohorts (table 1). There were minimal differences between the control populations. The coeliac cohorts were very similar in terms of male/female ratio, intestinal biopsy features at diagnosis, and *HLA-DQ* status. Data were not available to assess the proportion of cases with positive family history, a group in which the genetic contribution to disease susceptibility might be higher. Dutch coeliac cases had a lower median age at diagnosis (35 *v* 42 years) but overall there were only small differences in demographics between the populations.

Further genetic studies in different populations are now necessary to resolve whether *MYO9B* variants truly predispose to coeliac disease and, if so, obtain an accurate estimate of

effect size. Our large study provides a definitive result for the UK population. Functional studies may also illuminate the role of MYO9B and its putative role in influencing intestinal barrier function. Knowledge of *MYO9B* (and *HLA-DQ* variants) and analysis of potential gene-gene interactions might simplify identification of other coeliac disease susceptibility genes. The availability of new technologies enabling comparison of common human variation at a genome wide level makes this a realistic prospect.[20] Our understanding of coeliac disease pathogenesis seems set to increase rapidly.

ACKNOWLEDGMENTS

We thank all individuals who kindly donated blood samples. We thank Dr Andrew Walley (Imperial College Genome Centre), Dr Chaz Mein (Bart's and the London Genome Centre), Dr Martin Barnardo (Oxford Transplant Centre), and their colleagues for genotyping assistance. We acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. The study was funded by the Wellcome Trust (Clinician Scientist Fellowship to DAvH) and Coeliac UK.

Authors' affiliations

K A Hunt, R J Playford, D A van Heel, Institute of Cell and Molecular Science, Barts and The London, Queen Mary's School of Medicine and Dentistry, London, UK
J R F Walters, Division of Medicine, Imperial College London (Hammersmith Campus), London, UK
A J Monsuur, C Wijmenga, Complex Genetics Section, University Medical Centre Utrecht, the Netherlands
W L McArdle, Avon Longitudinal Study of Parents and Children, University of Bristol, Bristol, UK
P J Kumar, Department of Gastroenterology, Barts and the London NHS Trust, London, UK
S P L Travis, D P Jewell, Department of Gastroenterology, John Radcliffe Hospital, Oxford, UK
D P Strachan, Division of Community Health Sciences, St George's, University of London, London, UK

Conflict of interest: None declared.

REFERENCES

1 **van Heel DA**, Hunt K, Greco L, *et al*. Genetics in coeliac disease. *Best Pract Res Clin Gastroenterol* 2005;**19**:323–39.
2 **Greco L**, Romino R, Coto I, *et al*. The first large population based twin study of coeliac disease. *Gut* 2002;**50**:624–8.
3 **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.
4 **Monsuur AJ**, de Bakker PIW, Alizadeh BZ, *et al*. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. *Nat Genet* 2005;**12**:1341–4.
5 **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.
6 **Babron MC**, Nilsson S, Adamovic S, *et al*. Meta and pooled analysis of European coeliac disease data. *Eur J Hum Genet* 2003;**11**:828–34.
7 **Post PL**, Tyska MJ, O'Connell CB, *et al*. Myosin-IXb is a single-headed and processive motor. *J Biol Chem* 2002;**277**:11679–83.
8 **Muller RT**, Honnert U, Reinhard J, *et al*. 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.
9 **Russo JM**, Florian P, Shen L, *et al*. Distinct temporal-spatial roles for rho kinase and myosin light chain kinase in epithelial purse-string wound closure. *Gastroenterology* 2005;**128**:987–1001.
10 **Hunt KA**, McGovern DP, Kumar PJ, *et al*. A common CTLA4 haplotype associated with coeliac disease. *Eur J Hum Genet* 2005;**13**:440–4.
11 **Barrett JC**, Fry B, Maller J, *et al*. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;**21**:263–5.
12 **Qin ZS**, Niu T, Liu JS. Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am J Hum Genet* 2002;**71**:1242–7.
13 **Purcell S**, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;**19**:149–50.
14 **van Heel DA**, West J. Recent advances in coeliac disease. *Gut*. 2006;55: 000–0 (gt75119 in press)).
15 **Cardon LR**, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001;**2**:91–9.
16 **Economou M**, Trikalinos TA, Loizou KT, *et al*. Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. *Am J Gastroenterol* 2004;**99**:2393–404.
17 **Trinh TT**, Rioux JD. The promise and perils of interpreting genetic associations in Crohn's disease. *Gut* 2005;**54**:1354–7.
18 **Giallourakis C**, Stoll M, Miller K, *et al*. IBD5 is a general risk factor for inflammatory bowel disease: replication of association with Crohn disease and identification of a novel association with ulcerative colitis. *Am J Hum Genet* 2003;**73**:205–11.
19 **Negoro K**, McGovern DP, Kinouchi Y, *et al*. Analysis of the IBD5 locus and potential gene-gene interactions in Crohn's disease. *Gut* 2003;**52**:541–6.
20 **Altshuler D**, Brooks LD, Chakravarti A, *et al*. A haplotype map of the human genome. *Nature* 2005;**437**:1299–320.

Chapter 3.3

Association Analysis of *MYO9B* Gene Polymorphisms with Celiac Disease in a Swedish/Norwegian Cohort

Silja S. Amundsen, Alienke J. Monsuur, Martin C. Wapenaar, Benedicte A. Lie, Johan Ek, Audur H. Gudjónsdóttir, Henry Ascher, Cisca Wijmenga, and Ludvig M. Sollid

Human Immunology 2006; 67:341–345

Association Analysis of MYO9B Gene Polymorphisms with Celiac Disease in a Swedish/Norwegian Cohort

Silja S. Amundsen, Alienke J. Monsuur, Martin C. Wapenaar, Benedicte A. Lie, Johan Ek, Audur H. Gudjónsdóttir, Henry Ascher, Cisca Wijmenga, and Ludvig M. Sollid

ABSTRACT: Association between myosin IXB (MYO9B) gene variants and celiac disease (CD) has been reported in a study of a Dutch cohort. Six single nucleotide polymorphisms (SNPs) within the 3′ part of the MYO9B gene showed significant genetic association and formed an associated haplotype. The current study aimed to replicate these findings in a Swedish/Norwegian cohort. Genotyping of the three SNPs which tagged the associated haplotype was performed in a CD family dataset ($n = 326$) and in an additional set of healthy controls ($n = 562$). Although our material provided reasonable power to detect the previously observed association, we were unable to replicate association with these SNPs. Lack of reproducibility could be explained by no or negligible contribution of MYO9B to the genetic predisposition to CD in the Swedish/Norwegian population. Alternatively, it might be due to variable linkage disequilibria in distinct populations in the tested SNPs and a causative mutation yet to be identified or to false positive findings (type I error) in the Dutch study. *Human Immunology 67, 341–345 (2006).* © American Society for Histocompatibility and Immunogenetics, 2006. Published by Elsevier Inc.

KEYWORDS: Celiac disease; MYO9B; genetic association; myosin IXB

ABBREVIATIONS

CD	celiac disease
HLA	human leukocyte antigen
MYO9B	myosin IXB
SNP	single nucleotide polymorphism
LD	linkage disequilibrium

INTRODUCTION

Celiac disease (CD), or gluten-sensitive enteropathy, is a prevalent food intolerance disorder with a multifactorial etiology involving both environmental and genetic factors. CD is characterized by villous atrophy, crypt hyperplasia, leukocytic infiltration in the epithelium and lamina propria, and normalization of the mucosal alteration on a gluten-free diet [1]. Although categorized as a food hypersensitivity disorder, CD has autoimmune components best signified by antibodies to the self-antigen tissue transglutaminase (IgA TG) [2, 3]. Both HLA-linked and non-HLA-linked genes contribute to the genetic component. Several linkage screens have suggested some non-HLA candidate regions but, until recently, no susceptibility genes other than HLA-DQ have been determined.

Recently, myosin IXB (MYO9B) was identified as a susceptibility gene for CD in the Dutch population. Initially, a genomewide linkage screen demonstrated suggestive linkage to region 19p13.1 [4]. The linkage peak was found at marker D19S899 and the same marker demonstrated significant genetic association in an inde-

From the Institute of Immunology, University of Oslo, Rikshospitalet University Hospital, Oslo, Norway (S.S.A., L.M.S.); Institute of Immunology, Rikshospitalet-Radiumhospitalet Medical Center, Oslo, Norway (B.A.L.); Complex Genetics Section, DBG-Department of Medical Genetics, University Medical Center of Utrecht, Utrecht, the Netherlands (A.J.M., M.C.W., C.W.); Department of Pediatrics, Buskerud Hospital Trust, Drammen, Norway (J.E.); Department of Pediatrics, The Queen Silvia Children's Hospital, Göteborg University, Göteborg, Sweden (A.H.G., H.A.); and The Nordic School of Public Health, Göteborg, Sweden (H.A.).

Address reprint requests to: Silja Svanstrøm Amundsen, Institute of Immunology, University of Oslo, Rikshospitalet, N-0027 Oslo, Norway; Tel: +4723074227; Fax: +4723073510; E-mail: s.s.amundsen@medisin.uio.no.

Human Immunology **67,** 341–345 (2006)

0198-8859/06/$–see front matter
doi:10.1016/j.humimm.2006.03.020

pendent case–control cohort. The marker D19S899 is located in intron 1 of the MYO9B gene. In a follow-up study, six single nucleotide polymorphisms (SNPs) located within the 3′ part of MYO9B demonstrated highly significant genetic association with CD in the Dutch population both as single markers and as a haplotype [5]. The A allele at one of these SNPs, rs2305764, could alone explain the overall association observed at MYO9B. The risk of developing CD was increased in homozygous (odds ratio = 2.27) compared to heterozygous (odds ratio = 1.66) individuals for the A allele at rs2305764, thereby suggesting a codominant allele-dose effect model. Notably, linkage to 19p13.1 has not been replicated in other genomewide linkage studies of CD, except for one study where a weak linkage signal (Zlr = 1.65; $p = 0.05$) was seen when data from six European countries, including our Scandinavian material, were pooled together [6]. The importance of MYO9B in the etiology of CD would be reinforced if disease associations could be demonstrated also in other populations. The first, and to our knowledge the only, replication study to date reported lack of association of the MYO9B variants with CD in a British cohort [7].

MYO9B encodes a single motor protein that belongs to the class IX myosin family [8] and harbors a GAP domain in its tail for Rho family GTPase activation. Rho family GTPases are involved in several cellular processes including regulation of the actin cytoskeleton [9]. Rho-mediated regulation of the actin cytoskeleton is also of importance for the function of tight junctions and regulation of paracellular permeability [10, 11]. This latter fact therefore leads to the hypothesis that MYO9B is involved in the epithelial barrier function of the gut. An increased permeability of the gut epithelial lining in CD patients would allow the gluten peptides to enter the lamina propria more easily, thereby easing their way to the site of inflammation.

Our study attempted to verify the association found with the MYO9B gene in a Swedish/Norwegian population. We genotyped the three SNPs that tagged the MYO9B susceptibility haplotype in the Dutch study in 326 Swedish/Norwegian families. In addition, 562 healthy control subjects were genotyped and the individuals were, together with the total collection of family probands, used in a proband–control study.

MATERIAL AND METHODS

Patients and Controls

We used 326 CD families (100 multiplex and 226 simplex families previously described [12, 13]) from the relatively genetically homogeneous populations of Norway and Sweden. All CD patients fulfill the ESPGHAN (European Society of Pediatric Gastroenterology and Nutrition) diagnostic criteria [1], and all family members were of Swedish or Norwegian origin. The CD cases used in the proband–control study were unrelated probands selected from these families. The control group (n = 562) were recruited from the Norwegian Bone Marrow Donor Registry.

Genotyping

The three SNPs, rs2305767, rs1457092, and rs2305764, were genotyped using TaqMan by-design assays (Applied Biosystems, Foster City, CA, USA). The PCRs were run in 5-μl reaction volume using 1X Absolute QPCR Mix from ABgene, 0.5X TaqMan probe and primer mix, and genomic DNA (1–5 ng). Whole genome amplified DNA (5–20 ng) was used for the control samples (Genomiphi DNA amplification kit; Amersham Biosciences; amplified according to the manufacturers recommendation). PCRs were run in ABI PRISM 7000 and ABI PRISM 7900 machines (Applied Biosystems) using the following thermal profile: 95°C for 15 minutes and 45 cycles consisting of 95°C for 15 seconds and 60°C for 1 minute. Four negative and six positive controls were included on each plate. In addition, for each SNP assay, 180–192 randomly selected samples were run in duplicate to score the genotype error rate.

Statistical Analysis

The family data set was checked for Mendelian errors by using PedCheck [14]. For each SNP assay, 180–192 randomly selected samples were run in duplicate (this was done both for the family and for the control samples), and the genotyping error was estimated on the basis of the percentage of genotype calls deviating in the duplicates. The SNP genotypes were tested for deviation from Hardy–Weinberg equilibrium in the control samples, the family probands, and the genotypes of the parents separately.

Statistical power was calculated in our proband–control material prior to the study using the control frequencies and odds ratios (OR) obtained in the Dutch population. To calculate the statistical power to detect association with the rs2305764 A/A genotype, we used the power calculator located at http://pngu.mgh.harvard.edu/~purcell/gpc (significance level of 0.05 and OR = 1.66). The power to detect the susceptible AAA haplotype association was calculated using a two-sided χ^2 test of the power-calculator at http://calculators.stat.ucla.edu/powercalc (significance level of 0.05 and OR = 1.56).

Association analysis in families was performed by the transmission disequilibrium test (TDT) [15] employing the TDTPHASE application of the UNPHASED program [16]. For the proband–control study, allele frequencies were calculated using the COCAPHASE application in the UNPHASED program. For the family

TABLE 1 Association analyses of MYO9B SNPs in CD families and CD probands versus controls

	Family collection (n = 326)			Proband–Control (n = 326/562)			
SNP_Allele	T/NT	%T (CI)	p value	Proband (%)	Control (%)	OR (CI)	p value
rs2305767_A	149/147	50.3 (45–56)	0.91	395 (61.0)	609 (59.9)	1.04 (0.85–1.27)	0.68
rs1457092_A	161/148	52.1 (47–58)	0.46	246 (37.6)	296 (34.1)	1.17 (0.94–1.44)	0.16
rs2305764_A	164/156	51.3 (46–57)	0.66	273 (41.6)	452 (42.6)	0.96 (0.79–1.17)	0.68

Abbreviations: MYO9B = myosin IXB; SNPs = single nucleotide polymorphisms; CD = celiac disease; n = sample size; T/NT = number of transmitted versus nontransmitted alleles; %T = percentage of transmitted alleles; CI = 95% confidence interval; p value = uncorrected p value; Proband (%) = frequency of allele in probands; Control (%) = frequency of allele in controls; OR = odds ratio.

collection, the haplotype analysis included only certain haplotypes (using TDTPHASE), while for the proband–control study haplotype frequencies were estimated using the Expectation-Maximization (EM) algorithm implemented in the COCAPHASE program. Comparisons of the allele, genotype, and haplotype frequencies in patients and controls were performed using a χ^2 test (Woolf's formula) (only haplotypes with frequencies >5% were included in this analysis). A p value <0.05 was considered statistically significant. HaploView (version 3.2) [17] was used to determine linkage disequilibrium (LD) between the SNPs, and linkage was calculated using the Allegro (version 2.0) software [18].

RESULTS AND DISCUSSION

Prior to our study we estimated that our proband–control data set provided 99 and 86% power to detect an association with the rs2305764 A/A genotype and the susceptible AAA haplotype, respectively.

Pedcheck revealed Mendelian error rates of 1.5, 0.6, and 2.8% for rs2305767, rs1457092, and rs2305764, respectively, in the family material, whereas the samples run in duplicate revealed genotype error rates of 0, 1.1, and 1.6%, respectively. For the duplicated control samples, an error rate of 0% was seen for rs2305767 and rs1457092, whereas rs2305764 manifested an error rate of 3.3%. All SNPs were in Hardy–Weinberg equilibrium in both the control samples and the families (parents and probands separately).

First, we tested for association in the family material (n = 326) using TDT. We found no support for single-point association with any of the three tag SNPs (Table 1). In an attempt to further increase the power, we included an additional set of independent controls and switched to a case–control setting. In total, 570 controls were analyzed together with the 326 CD probands, extracted from the families, as cases. However, no single-point association was observed for the proband–control study either (Table 1).

Second, we performed haplotype analysis combining the three SNPs rs2305767, rs1457092, and rs2305764. No significant association for the previously reported AAA susceptibility haplotype was seen (Table 2). We also determined LD between the SNPs and found strong LD between the three SNPs in both cases ($D' > 0.95$) and controls ($D' > 0.90$). This is consistent with the LD structure observed in the Dutch study.

Since the strongest association at rs2305764 had been reported for homozygous A/A (OR = 2.27) compared to heterozygous A/G individuals (OR = 1.66), we tested this dose effect model using χ^2 statistics. In our material, however, no signs of dose effect were evident, because the A/A genotype did not display any association (OR = 0.94; 95% CI = 0.66–1.33; p = 0.72).

To investigate whether we could find any evidence for linkage to this region independent of association, we performed nonparametric linkage analyses in our 100 multiplex families. No convincing evidence for linkage was observed by multipoint analyses (max NPL = 0.5)

TABLE 2 Association analyses of MYO9B haplotypes in CD families and CD probands versus controls

	Family collection (n = 326)				Proband–Control (n = 326/562)			
Haplotype	T/NT	RR	%T (CI)	p value	Proband (%)	Control (%)	OR (CI)	p value
GCG	98/96	1.00	50.5 (43–58)	0.90	250 (38.8)	424 (38.3)	1.00 (0.81–1.21)	0.95
ACG	68/87	0.78	43.9 (36–52)	0.15	120 (18.7)	191 (17.3)	1.07 (0.83–1.38)	0.40
ACA	23/18	1.25	56.1 (41–71)	0.43	30 (4.7)	83 (7.5)	0.60 (0.39–0.91)	0.01
AAA	112/94	1.05	54.4 (48–61)	0.28	236 (36.7)	380 (34.3)	1.07 (0.87–1.31)	0.31

Abbreviations: MYO9B = myosin IXB; CD = celiac disease; n = sample size; T/NT = number of transmitted versus nontransmitted haplotypes; RR = relative risk; %T = percentage of transmitted haplotypes; CI = 95% confidence interval; p value = uncorrected p value; Proband (%) = frequencies of haplotypes in probands; Controls (%) = frequencies of haplotypes in controls. Global p value for the haplotype test of the family study was 0.088 and for the proband–control study 0.002. Four common haplotypes (frequencies >5%) were present in both test settings and are presented.

nor by using the SNP haplotypes as a composite marker (NPL = 1.1).

Assuming the same OR and allele frequencies as obtained in the Dutch study, our sample size should provide sufficient power to detect an association with both the rs2305764 A/A genotype and the susceptible AAA haplotype. Despite reasonable power, we were unable to detect association between the three tested MYO9B SNPs and CD in our Swedish/Norwegian cohort. The susceptibility effect exercised by MYO9B may vary between populations, which would influence the power of our material to observe an association. It should also be taken into account that initial association reports often overestimate disease risk [18]. Therefore, if the true OR is lower and, in the worst case, as low as the lower confidence interval, power to detect the effect in our material decreases dramatically from >95 to 61%. This could partly explain why we were unable to replicate the Dutch findings.

The discrepant results might also indicate that there are population-specific effects of MYO9B on CD susceptibility with no or negligible contribution of MYO9B in the Swedish/Norwegian population. Alternatively, MYO9B may be a CD susceptibility gene also in the Swedish/Norwegian population, but the causative SNP has variable LD to the hitherto tested markers in different populations.

Finally, a possible explanation would be that the initial Dutch report was a false positive finding due to type I error. The study of Monsuur *et al.* [5] involved multiple testing of a large number of genetic markers. Although the association signal withstood conservative Bonferroni correction, more reports testing MYO9B variants in independent populations are needed to be able to draw conclusions about the role of MYO9B as a susceptibility gene in CD.

Acknowledgments

This work has been supported by the Research Council of Norway, by grants for the Netherlands Organization for Scientific Research (Grants 912-02-028) and the Celiac Disease Consortium, an innovative cluster approved by the Netherlands Genomics Initiative and partially founded by the Dutch government (Grant BSIK03009). We thank the families participating in the study and Britt-Marie Käck and the Swedish Coeliac Society for help collecting families and blood samples. The Norwegian Bone Marrow Donor Registry is acknowledged for the collection of control samples.

REFERENCES

1. Walker-Smith J, Guandalini S, Schmitz J, Schmerling D, Visakorpi J: Report of working group of European Society of Paediatric Gastroenterology and Nutrition: revised criteria for diagnosis of coeliac disease. Arch Dis Child 65:909, 1990.
2. Dieterich W, Ehnis T, Bauer M, Donner P, Volta U, Riecken EO, Schuppan D: Identification of tissue transglutaminase as the autoantigen of celiac disease. Nat Med 3:797, 1997.
3. Sollid LM: Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2:647, 2002.
4. van Belzen MJ, Meijer JW, Sandkuijl LA, Bardoel AF, Mulder CJ, Pearson PL, Houwen RH, Wijmenga C: A major non-HLA locus in celiac disease maps to chromosome 19. Gastroenterology 125:1032, 2003.
5. Monsuur AJ, Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, Franke L, Slot RV, Belzen MJ, Lavrijsen IC, Diosdado B, Daly MJ, Mulder CJ, Mearin ML, Meijer JW, Meijer GA, Oort E, Wapenaar MC, Koeleman BP, Wijmenga C: Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. Nat Genet 37:1341, 2005.
6. Babron MC, Nilsson S, Adamovic S, Naluai AT, Wahlstrom J, Ascher H, Ciclitira PJ, Sollid LM, Partanen J, Greco L, Clerget-Darpoux F: Meta and pooled analysis of European coeliac disease data. Eur J Hum Genet 11:828, 2003.
7. Hunt KA, Monsuur AJ, McArdle W, Kumar PJ, Travis SP, Walters JR, Jewell DP, Strachan DP, Playford RJ, Wijmenga C, van Heel DA: Lack of association of MYO9B genetic variants with coeliac disease in a British cohort. Gut 2006.
8. Post PL, Tyska MJ, O'Connell CB, Johung K, Hayward A, Mooseker MS: Myosin-IXb is a single-headed and processive motor. J Biol Chem 277:11679, 2002.
9. Sahai E, Marshall CJ: RHO-GTPases and cancer. Nat Rev Cancer 2:133, 2002.
10. Bruewer M, Hopkins AM, Hobert ME, Nusrat A, Madara JL: RhoA, Rac1, and Cdc42 exert distinct effects on epithelial barrier via selective structural and biochemical modulation of junctional proteins and F-actin. Am J Physiol Cell Physiol 287:C327, 2004.
11. Matter K, Balda MS: Signalling to and from tight junctions. Nat Rev Mol Cell Biol 4:225, 2003.
12. Gudjonsdottir AH, Nilsson S, Ek J, Kristiansson B, Ascher H: The risk of celiac disease in 107 families with at least two affected siblings. J Pediatr Gastroenterol Nutr 38:338, 2004.
13. Louka AS, Nilsson S, Olsson M, Talseth B, Lie BA, Ek J, Gudjonsdottir AH, Ascher H, Sollid LM: 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 60:147, 2002.
14. O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259, 1998.

15. 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 52:506, 1993.
16. Dudbridge F: Pedigree disequilibrium tests for multilocus haplotypes. Genet Epidemiol 25:115, 2003.
17. Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263, 2005.
18. Colhoun HM, McKeigue PM, Davey SG: Problems of reporting genetic associations with complex outcomes. Lancet 361:865, 2003.

Chapter 3.4

Genetic variation in *myosin IXB* is associated with ulcerative colitis

Alienke J. Monsuur, Adriaan A. van Bodegraven, Christine R. Curley, Karen A. Hunt, Ronald K. Linskens, Clive M. Onnie, J. Bart A. Crusius, Vito Annese, Anna Latiano, Mark S. Silverberg, Alain Bitton, Sheila A. Fisher, A. Hilary Steinhart, Alastair Forbes, Jeremy Sanderson, Natalie J. Prescott, David P. Strachan, Raymond J. Playford, Christopher G. Mathew, Cisca Wijmenga, Mark J. Daly, John D. Rioux, David A. van Heel

Gastroenterology 2006; 131(6):1768-1774

BASIC–ALIMENTARY TRACT

Genetic Variation in Myosin IXB Is Associated With Ulcerative Colitis

ADRIAAN A. VAN BODEGRAVEN,[*] CHRISTINE R. CURLEY,[‡] KAREN A. HUNT,[§] ALIENKE J. MONSUUR,[‖] RONALD K. LINSKENS,[¶] CLIVE M. ONNIE,[#] J. BART A. CRUSIUS,[**] VITO ANNESE,[‡‡] ANNA LATIANO,[‡‡] MARK S. SILVERBERG,[§§] ALAIN BITTON,[‖‖] SHEILA A. FISHER,[#] A. HILARY STEINHART,[§§] ALASTAIR FORBES,[¶¶] JEREMY SANDERSON,[##] NATALIE J. PRESCOTT,[#] DAVID P. STRACHAN,[***] RAYMOND J. PLAYFORD,[§] CHRISTOPHER G. MATHEW,[#] CISCA WIJMENGA,[‖] MARK J. DALY,[‡,‡‡‡] JOHN D. RIOUX,[‡,§§§] and DAVID A. VAN HEEL[§]

[*]Department of Gastroenterology, VU University Medical Centre, Amsterdam, The Netherlands; [‡]The Broad Institute, Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts; [§]Institute of Cell and Molecular Science, Barts & The London, Queen Mary's School of Medicine & Dentistry, London, United Kingdom; [‖]DBG-Department of Medical Genetics, University Medical Center Utrecht, CG Utrecht, The Netherlands; [¶]Department of Gastroenterology and Internal Medicine, St Anna Ziekenhuis, Geldrop, The Netherlands; [#]Department of Medical and Molecular Genetics, King's College London School of Medicine, Guy's Hospital, London, United Kingdom; [**]Laboratory of Immunogenetics, Department of Pathology, VU University Medical Centre, Amsterdam, The Netherlands; [‡‡]CSS-IRCCS Hospital, San Giovanni Rotondo, Italy; [§§]Sinai Hospital Inflammatory Bowel Disease Centre, University of Toronto, Toronto, Ontario, Canada; [‖‖]Department of Gastroenterology, McGill University Health Centre, Royal Victoria Hospital, Montréal, Québec, Canada; [¶¶]Department of Gastroenterology, University College London Hospitals, London, United Kingdom; [##]Department of Gastroenterology, Guy's Hospital, London, United Kingdom; [***]Division of Community Health Sciences, St George's Hospital Medical School, London, United Kingdom; [‡‡‡]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts; and the [§§§]Université de Montréal, Montreal Heart Institute, Montréal, Québec, Canada

Background & Aims: Common germline genetic variation in the 3′ region of myosin IXB (*MYO9B*) has been associated recently with susceptibility to celiac disease, with a hypothesis that *MYO9B* variants might influence intestinal permeability. These findings suggested the current study investigating a possible further role for *MYO9B* variation in inflammatory bowel disease. **Methods:** Eight single-nucleotide polymorphisms (SNPs) were selected to tag common haplotypes from the 35-kb 3′ region of *MYO9B*. These included the strongest celiac disease–associated variants reported in a Dutch cohort. These SNPs were studied in 3 independently collected and genotyped case-control cohorts of European descent (UK, Dutch, and Canadian/Italian), comprising in total 2717 inflammatory bowel disease patients (1197 with Crohn's disease, 1520 with ulcerative colitis) and 4440 controls. **Results:** Common variation in *MYO9B* was associated with susceptibility to inflammatory bowel disease in all 3 cohorts examined (most associated SNP, rs1545620; meta-analysis $P = 1.9 \times 10^{-6}$; odds ratio, 1.2), with the same alleles showing association as reported for celiac disease. **Conclusions:** MYO9B genetic variants predispose to inflammatory bowel disease. Interestingly, rs1545620 is a nonsynonymous variant leading to an amino acid change (Ala1011Ser) in the third calmodulin binding IQ domain of MYO9B. Unlike previous variants (in other genes) reported to predispose to inflammatory bowel disease, the association at *MYO9B* was considerably stronger with ulcerative colitis, although weaker association with Crohn's disease also was observed. These data imply shared causal mechanisms underlying intestinal inflammatory diseases.

The inflammatory bowel diseases (IBDs)—clinically classified as Crohn's disease or ulcerative colitis—are common chronic disorders of the gastrointestinal tract with increasing incidence in urbanized developed countries. At least 9 broad genomic regions (susceptibility loci) have been implicated in IBD via linkage analysis,[1–3] but from these only 2 compelling association findings have been validated repeatedly: variants in *CARD15* (also known as *NOD2*) on chromosome 16, and variants throughout a long (~250 kb) region of chromosome 5q31 (including a coding variant in *SLC22A4*).[2] In both of these cases the association appears predominantly or exclusively to Crohn's disease. Although there is some evidence regarding the HLA region and *CARD4* (*NOD1*),[3,4] there has been little compelling association evidence presented to date for ulcerative colitis.

Recently, genetic variation in the 3′ region of the myosin IXB (*MYO9B*) gene was reported to be associated with celiac disease.[5] This gene was identified from the chromosome 19p13 *CELIAC4* linkage locus by comprehensive screening of the region using 291 single-nucleotide polymorphisms (SNPs) across the peak 6-Mb linkage region in a cohort of Dutch celiac disease patients and controls.[5] These tag SNPs were selected to capture most of the common genetic variation present in this region, by virtue of being strongly correlated with most other known variants.[6,7] The 19p13 region also has been linked to IBD (*IBD6*).[1,8] In the Dutch celiac study, significant and replicable association with celiac disease ($P = 2.1 \times 10^{-6}$) was found for a common variant (rs2305764) in intron 28 of *MYO9B*.[5] Heterozygotes and homozygotes for the A allele of rs2305764 had increased celiac disease risk by 1.7-fold and 2.3-fold, respectively. Another nearby correlated variant, rs1457092 ($r^2 = .80$ to rs2305764), was also significantly associated in this study, and was not convincingly excluded as potentially causal.[3] Over expression of rat myosin IXB leads to actin filament–related morphologic changes in epithelial cells,[9] and human myosin IXB is expressed in intestinal epithelial cells.[10] Although the mechanism by which the *MYO9B* gene variants associated with celiac disease lead to altered function in the gut and subsequent celiac disease

Abbreviations used in this paper: MYO9B, myosin IXB; SNP, single-nucleotide polymorphism.

0016-5085/06/$32.00
doi:10.1053/j.gastro.2006.09.011

Table 1. Demographics of Case-Control Cohorts

		No. of individuals	Median age, *y*	Age range	Median age at diagnosis	Age range at diagnosis	% female
Dutch	Crohn's disease	298	44	17–85	33	15–77	70.5
	Ulcerative colitis	290	48	19–89	39	15–86	44.6
	Controls	1624[a]	47	6–93			43.5
UK[b]	Crohn's disease	735	41	17–89	23	0–77	57.7
	Ulcerative colitis	580	51	21–94	28	0–81	50.3
	Controls	2371	48	23–65	—	—	49.7
Canadian/Italian	Crohn's disease	164	44	19–76	23	13–71	35.4
	Ulcerative colitis	650	45	18–83	34	3–75	39.4
	Controls	445	43	20–75			17.7

[a]Age and sex data available for 1149 individuals.
[b]Age data available on 630 Crohn's disease cases, 460 ulcerative colitis cases, and 2045 controls; age at diagnosis data available on 545 Crohn's disease cases, 449 ulcerative colitis cases; and sex data available for 728 Crohn's disease cases, 574 ulcerative colitis cases, and 2132 controls.

pathogenesis currently is unclear, we hypothesized that *MYO9B* variants also might promote susceptibility to other intestinal inflammatory diseases, including Crohn's disease and ulcerative colitis.

To investigate this hypothesis we tested genetic variants in the 3′ region of *MYO9B* for association with IBD because this region contains all SNPs associated with celiac disease in the Dutch population,[5] including the 2 SNPs (rs1457092 and rs2305764) reported to show the strongest association.

Materials and Methods

Cohorts

We genotyped large samples of Crohn's disease and ulcerative colitis cases and controls. All individuals studied were of European Caucasian descent. Informed consent was obtained from all subjects. Cases were diagnosed according to established clinical criteria (including review of clinical notes, radiologic, histologic, and endoscopic data). Ethical approval for the study was obtained from the following: Dutch cohorts: VU University Medical Center Amsterdam Ethics Committee; UK cohorts: SouthEast Multicentre Research Ethics Committee, Oxford Research Ethics Committee, Guy's Hospital Research Ethics Committee, Harrow Research Ethics Committee; and Canadian/Italian cohorts: San Giovanni Rotondo Hospital Ethics Committee and the University of Toronto Ethics Committee.

IBD case and control cohorts are summarized in Table 1. Dutch cases were recruited from the Vrije Universiteit Medical Center in Amsterdam.[11] British cases were collected from St. Thomas' Hospitals (London), St. Mark's Hospital (London), and King's College Hospital (London).[12–14] The Canadian (n = 38) and Southern Italian case (n = 776) cohorts were collected from multiple sites in the province of Quebec, Canada, and from San Giovanni Rotondo CSS Hospital in Italy.[13,14] The Dutch control samples comprise the random hospital controls (n = 220) and blood bank donors (n = 470) previously described,[5] additionally supplemented with Dutch blood bank donors (n = 934). British population controls were drawn from the 1958 British Birth Cohort (n = 2000), and healthy controls from blood bank donors and institutional staff (n = 371). Canadian (n = 57) and Southern Italian (n = 388) controls were recruited from healthy blood donors. No significant difference in *MYO9B* allele frequencies was observed between the UK population and healthy controls, or between Canadian and Italian controls (data not shown). All control genotypes were in Hardy–Weinberg equilibrium (data not shown, $P > .05$).

SNP Selection and Experimental Design

We desired to comprehensively assess the 3′ region of *MYO9B* for association with IBD. All 5 variants (rs2305767, rs962917, rs2279003, rs2305765, and rs2305764) showing replicated association with celiac disease[5] (in 2 independent cohorts from the Dutch population) are contained within a 34.5-kb region around the 3′ region of the *MYO9B* gene (Figure 1). This region also includes the 2 SNPs (rs1457092 and rs2305764) showing the strongest association in this report,[5] neither of which was necessarily the unique celiac disease variant. We selected these and any other nearby SNPs that were highly correlated (in the European data set from the HapMap project[7]) and also therefore might be potential disease variant candidates. We did not select genetically redundant SNPs showing complete correlation ($r^2 = 1$).

We chose 8 SNPs for genotyping that efficiently tag (ie, strongly correlate at $r^2 > .8$ with all other variants, including the 6 earlier-described unique SNPs from the Dutch celiac study[5]) the 3′ region of *MYO9B* (Figure 1). This region spans 34.5 kb from rs7259292 to rs388484, and contains in total 22 polymorphic phase I/II HapMap SNPs (European origin CEU population; HapMap data release 20, January 2006).[7] Analysis using Haploview[15] showed this region to contain 2 haplotype blocks (regions of genomic DNA sequence showing strong correlation between variants, using Gabriel et al[16] criteria). In particular, this region also contains all SNPs in the phase I/II HapMap data set[7] that are partially correlated ($r^2 > .5$) to these associated variants in the celiac study (data not shown).

Genotyping

Each of the 3 laboratories (Barts & The London, University Medical Centre Utrecht, and Broad Institute) performed genotyping independently. Genotyping methods used were as follows: Dutch cohort (Taqman assay; Applied Biosystems, Foster City, CA), Canadian and Italian cohorts (Sequenom assay; Sequenom, San Diego, CA), British cohorts (Taqman assay, with identical genotypes obtained by Sequenom assay on a

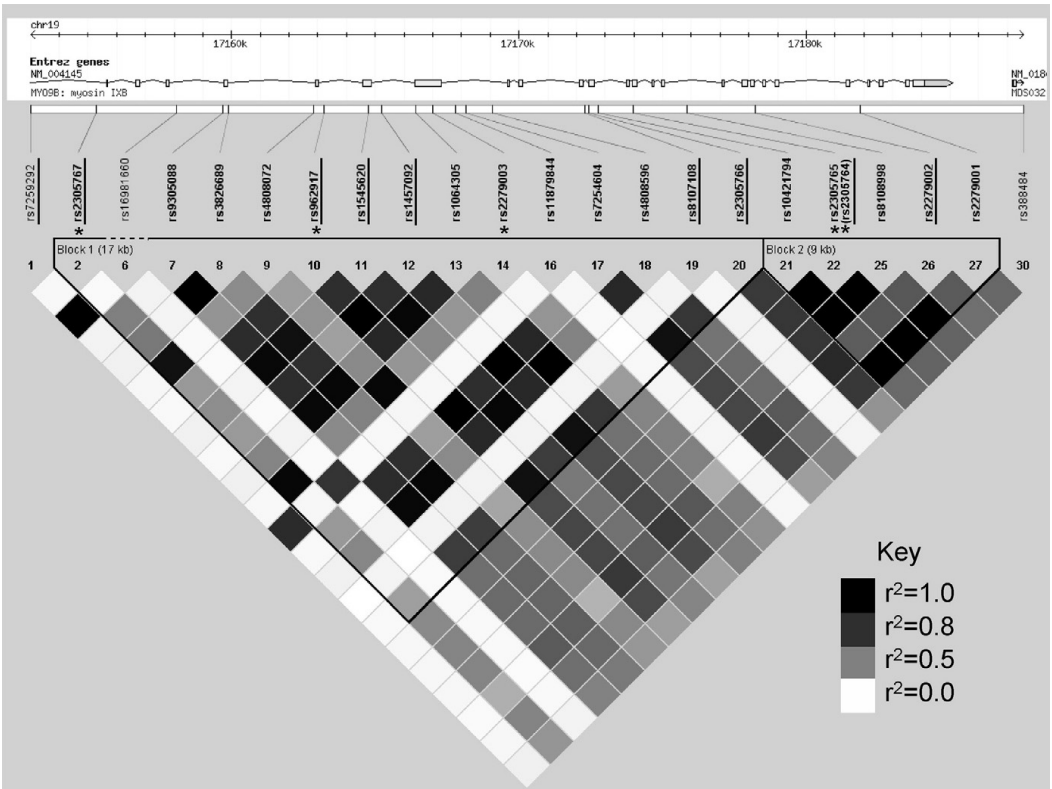


Figure 1. SNP selection from the 3′ region of MYO9B. HapMap SNPs and Haploview linkage disequilibrium analysis shown from the 34.5-kb 3′ region of MYO9B. Greyscale boxes indicate correlation coefficient between variants. *Five SNPs reported to show replicated association with celiac disease in 2 cohorts from the Dutch population.[5] Underlined SNPs were genotyped in the current study. SNP rs2305764 was chosen for the current study, however, this was not genotyped in the HapMap project but was correlated almost completely with the adjacent (841-bp apart) HapMap SNP rs2305765 in the UK data set (r^2 = .98, n = 2371 controls). Dutch and UK cohorts were genotyped for rs1457092, Canadian/Italian cohorts for rs962917—these variants show almost complete correlation (see Materials and Methods section).

subset of 342 DNA samples[17]). Full details of primers and probes are available on request.

SNP rs1457092 failed for technical reasons at the site typing the Canadian/Italian cohort. This cohort was therefore evaluated for neighboring SNP rs962917 because these 2 variants are less than 2-kb apart, and are near-perfectly correlated (r^2 = 1.00, in the HapMap European origin CEU data set[7]; r^2 = .98 in 686 Dutch controls in the earlier celiac study[5]). SNP rs2305766 failed for technical reasons at the site typing the UK sample.

The overall genotype success rate was 96.5% (range, 91.5%–99.5% observed per SNP per population group). Statistical analysis was performed using 2-tailed χ^2 tests of case vs control allele counts in Haploview v3.2.[15] M.J.D. performed the final analyses on all data sets.

Results

In total, 1197 Crohn's disease cases, 1520 ulcerative colitis cases, and 4440 controls were genotyped with 8 tagging SNPs across the celiac disease–associated region of *MYO9B* in the current study, in which the sample assembly, genotyping, and initial analysis were performed completely independently at 3 sites (Table 1).

Significant association with IBD across all 3 cohorts was shown for several common variants in this 3′ region, including the same celiac-associated alleles of SNPs rs1457092 and rs2305764 (Table 2). A meta-analysis (using the Mantel–Haenszel method) of these data identified the strongest associations to rs1545620 ($P = 1.9 \times 10^{-6}$), with an estimated odds ratio of 1.19 (95% confidence interval, 1.11–1.28) and to rs1457092 ($P = 2.1 \times 10^{-6}$; odds ratio, 1.20; 95% confidence interval, 1.11–1.29). The data imply a log-additive model—no apparent dominant or recessive effects are detectable.

We then analyzed the 2 subphenotypes separately for the maximally associated SNPs and found that, although association is seen to both Crohn's disease and ulcerative colitis, the association is particularly strong in the ulcerative colitis subset

Table 2. Association Analysis of *MYO9B* SNPs in IBD

	IBD			Controls				
	Allele counts			Allele counts				
	Major	Minor	Minor allele frequency	Major	Minor	Minor allele frequency	P^a by cohort	P^b meta-analysis
rs7259292 (intron 14)								
Dutch	1133	31	0.027	3119	91	0.028	.76	
UK	2387	81	0.033	4359	155	0.034	.74	Not done
Canadian/Italian	1402	90	0.060	764	48	0.059	.91	
rs2305767 (intron 14)								
Dutch	705	437	0.383	1764	1392	0.441	.0006	.0020
UK	1515	1015	0.401	2694	1900	0.414	.31	
Canadian/Italian	963	631	0.396	501	365	0.421	.22	
rs1545620 (exon 20)								
Dutch	687	473	0.408	2056	1148	0.358	.0028[c]	.0000019
UK	1412	1112	0.441	2748	1848	0.402	.0016[c]	
Canadian/Italian	842	724	0.462	477	341	0.417	.034[c]	
rs1457092 (Dutch/UK)/rs962917 (Can/It) (intron 20)								
Dutch	708	438	0.382	2120	1042	0.330	.0013	.0000021
UK	1413	975	0.408	2889	1689	0.369	.0013	
Canadian/Italian	966	654	0.404	559	327	0.369	.090	
rs8107108 (intron 24)								
Dutch	1058	90	0.078	2959	245	0.076	.83	
UK	2319	163	0.066	4259	355	0.077	.082	Not done
Canadian/Italian	Not genotyped							
rs2305766 (intron 24)								
Dutch	731	437	0.374	2114	1050	0.332	.0093	
UK	Not genotyped							Not done
Canadian/Italian	944	628	0.399	524	310	0.372	.18	
rs2305764 (intron 28)								
Dutch	638	474	0.426	1958	1204	0.381	.0075	.00014
UK	1436	1126	0.440	2728	1882	0.408	.010	
Canadian/Italian	907	649	0.417	526	338	0.391	.21	
rs2279002 (intron 32)								
Dutch	758	382	0.335	2263	933	0.292	.0065	.00042
UK	1620	860	0.347	3161	1457	0.316	.0074	
Canadian/Italian	1071	471	0.305	565	241	0.299	.75	

[a]Two-tailed *P* values calculated by χ^2 analysis of allele counts.
[b]Mantel–Haenzsel method.
[c]*P* values also were obtained for rs1545620 by 100,000 random permutations of affection status in each of the 3 cohorts and were similar (Dutch, $P = .0031$; UK, $P = .0014$; Canadian/Italian, $P = .030$).

(Table 3) (meta-analysis results: rs1545620: $P = 1.3 \times 10^{-5}$; odds ratio, 1.23; 95% confidence interval, 1.12–1.35; rs1457092: $P = 7.8 \times 10^{-6}$; odds ratio, 1.24; 95% confidence interval, 1.13–1.36). More detailed phenotype–genotype analysis in the UK cohort (data not shown) did not reveal stronger association in any further subphenotype tested (age at diagnosis and disease location).

Conclusions

We observed significant association with *MYO9B* genetic variants for IBD susceptibility in 3 independently genotyped and collected cohorts. Interestingly, the most associated SNP rs1545620 induces a coding change (Ala1011Ser, Swiss-Prot Q13459) within the third IQ domain of MYO9B. The neck region of the MYO9B protein, between the motor and tail domains, contains 4 of these widely conserved 25 amino acid sequences that each form an α-helix capable of binding calmodulin.[18,19] Calmodulin regulates the processive motor activity of MYO9B on actin filaments in a Ca^{2+}-dependent manner. It is therefore plausible that rs1545620 might be the actual disease-causing variant influencing the velocity of the MYO9B protein. These results, taken together with our hypothesis-driven high prior probability of association, lead us to confidently propose this locus as associated with IBD.

The prior probability of detecting association for a multifactorial disease, such as IBD, is low.[20] However, several features of the current study add weight to the findings: (1) *MYO9B* variants are associated with another intestinal inflammatory disorder (celiac disease); (2) association was seen in 3 independent cohorts; and (3) a coding variant was identified. The significance level obtained by analysis of large cohorts in the current study also approaches the range suggested for genome-wide significance (ie, *P* values of 10^{-6} to 10^{-7}, allowing for correction owing to testing of variation in all ~30,000 human genes).[21–23] Permutation analyses on the genotype data sets to account for

Table 3. Association Analysis of *MYO9B* SNPs in Crohn's Disease and Ulcerative Colitis Subphenotypes

	Crohn's disease			Ulcerative colitis			Controls				
	Allele counts			Allele counts			Allele counts			P^a	
	Major	Minor	Minor allele frequency	Major	Minor	Minor allele frequency	Major	Minor	Minor allele frequency	Crohn's disease	*P* ulcerative colitis
rs7259292 (intron 14)											
Dutch	568	18	0.031	565	13	0.022	3119	91	0.028	.75	.43
UK	1347	49	0.035	1040	32	0.030	4359	155	0.034	.89	.46
Canadian/Italian	257	15	0.055	1145	75	0.061	764	48	0.059	.81	.83
rs2305767 (intron 14)											
Dutch	347	233	0.402	358	204	0.363	1764	1392	0.441	.079	.0006
UK	829	589	0.415	686	426	0.383	2694	1900	0.414	.90	.063
Canadian/Italian	187	137	0.423	776	494	0.389	501	365	0.421	.97	.13
rs1545620 (exon 20)											
Dutch	345	239	0.409	342	234	0.406	2056	1148	0.358	.019	.028
UK	805	615	0.433	607	497	0.450	2748	1848	0.402	.038	.0035
Canadian/Italian	180	134	0.427	662	590	0.471	477	341	0.417	.763	.015
rs1457092 (Dutch/UK) / rs962917 (Can/It) (intron 20)											
Dutch	358	222	0.383	350	216	0.382	2120	1042	0.330	.013	.016
UK	827	541	0.395	586	434	0.425	2889	1689	0.369	.075	.0008
Canadian/Italian	201	123	0.380	765	531	0.410	559	327	0.369	.74	.056
rs8107108 (intron 24)											
Dutch	536	44	0.076	522	46	0.081	2959	245	0.076	.96	.71
UK	1289	99	0.071	1030	64	0.059	4259	355	0.077	.49	.036
Canadian/Italian	Not genotyped										
rs2305766 (intron 24)											
Dutch	370	220	0.373	361	217	0.375	2114	1050	0.332	.053	.042
UK	Not genotyped										
Canadian/Italian	201	113	0.360	743	515	0.409	524	310	0.372	.71	.084
rs2305764 (intron 28)											
Dutch	333	233	0.412	305	241	0.441	1958	1204	0.381	.16	.0073
UK	814	606	0.427	622	520	0.455	2728	1882	0.408	.22	.0039
Canadian/Italian	191	119	0.384	716	530	0.425	526	338	0.391	.82	.12
rs2279002 (intron 32)											
Dutch	379	197	0.342	379	185	0.328	2263	933	0.292	.016	.084
UK	910	468	0.340	710	392	0.356	3161	1457	0.316	.092	.010
Canadian/Italian	229	87	0.275	842	384	0.313	565	241	0.299	.43	.50

[a]Two-tailed *P* values calculated by χ^2 analysis of allele counts.

the multiple testing of 8 SNPs leads to a *P* value correction of 4–6 times.

The evidence for association of MYO9B genetic variation with IBD was strongest in the Dutch and UK populations, although broadly similar control allele frequencies were observed across all 3 cohorts. The relatively weaker support in the Canadian/Italian cohorts might possibly be explained by a greater proportion of individuals of Southern vs Northern European extraction and a difference in susceptibility factors (either genetic or environmental) between these groups. Strong evidence for association remained in a repeat analyses after removing the Canadian/Italian cohort (rs1457092 meta-analysis: $P = 8.7 \times 10^{-6}$; rs1545620: $P = 1.9 \times 10^{-5}$). In addition to attempting to confirm the current findings, further studies should investigate the effect of ethnic origin on disease association.

Although chromosome 19 has been linked to IBD (*IBD6*),[1] we note that the odds ratio and frequency of the risk allele described here implies a low contribution to sibling risk ($\lambda s < 1.02$) and thus does not substantively contribute to the previous observations of linkage.

The observation that inherited variation in *MYO9B* predisposes to both celiac disease and ulcerative colitis is intriguing, and underscores current hypotheses that common cellular and molecular pathogenic mechanisms might be involved in multiple inflammatory diseases. The general observation of shared genetic architecture among other inflammatory and autoimmune diseases (eg,*CTLA4* and *PTPN22*) suggests that it may be worthwhile to evaluate the role of *MYO9B* in other conditions. It is interesting that although inflammation in celiac disease affects the proximal small bowel, inflammation in ulcerative colitis involves the large bowel. It is possible that *MYO9B* variants generally influence intestinal function, and that specific disease type is determined by other genetic and environmental factors.

It is not clear, at present, how genetic variation affects cellular mechanisms involving *MYO9B*. Human myosin IXB is expressed most strongly in leukocytes. It also is expressed in several other tissues and cell types—including intestinal and other epithelial cell lines.[10] MYO9B has processive motor properties, potentially switchable by regulatory mechanisms.[24,25] Unusually, MYO9B also has a Rho-guanosine triphosphatase activation domain that can negatively control Rho proteins. Rho kinases are involved in cytoskeletal modifications and affect tight junction assembly,[26] suggesting one hypothesis whereby *MYO9B* variants influence intestinal permeability lead-

ing to inflammatory intestinal disease susceptibility. *MYO9B* also is involved in control of bacterial invasion of epithelial cells, suggesting a second possible disease mechanism.[27] Finally, activated Rho proteins are involved in purse-string closure of human intestinal epithelial wounds, suggesting a role for MYO9B in intestinal restitution and repair after injury.[28] Further studies are now necessary to determine how *MYO9B* variants influence the shared causal mechanisms underlying intestinal inflammatory diseases, and whether these variants predispose more generally to other inflammatory and autoimmune diseases.

References

1. van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD, Lewis CM. Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. Hum Mol Genet 2004;13:763–770.
2. Schreiber S, Rosenstiel P, Albrecht M, Hampe J, Krawczak M. Genetics of Crohn disease, an archetypal inflammatory barrier disease. Nat Rev Genet 2005;6:376–388.
3. Mathew CG, Lewis CM. Genetics of inflammatory bowel disease: progress and prospects. Hum Mol Genet 2004;13:R161–R168.
4. McGovern DP, Hysi P, Ahmad T, van Heel DA, Moffatt MF, Carey A, Cookson WO, Jewell DP. Association between a complex insertion/deletion polymorphism in NOD1 (CARD4) and susceptibility to inflammatory bowel disease. Hum Mol Genet 2005;14: 1245–1250.
5. Monsuur AJ, de Bakker PIW, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, Franke L, van't Slot R, van Belzen MJ, Lavrijsen ICM, Diosdado B, Daly MJ, Mulder CJJ, Mearin ML, Meijer JWR, Meijer GA, van Oort E, Wapenaar MC, Koeleman BPC, Wijmenga C. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. Nat Genet 2005; 12:1341–1344.
6. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. Nat Genet 2001;29:233–237.
7. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. Nature 2005; 437:1299–1320.
8. Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, McLeod RS, Griffiths AM, Green T, Brettin TS, Stone V, Bull SB, Bitton A, Williams CN, Greenberg GR, Cohen Z, Lander ES, Hudson TJ, Siminovitch KA. Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. Am J Hum Genet 2000;66:1863–1870.
9. 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;8:2039–2053.
10. 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;109:653–661.
11. Xia B, Crusius JB, Wu J, Zwiers A, van Bodegraven AA, Pena AS. Signal transducer and activator of transcription 6 gene G2964A polymorphism and inflammatory bowel disease. Clin Exp Immunol 2003;131:446–450.
12. Cuthbert AP, Fisher SA, Mirza MM, King K, Hampe J, Croucher PJ, Mascheretti S, Sanderson J, Forbes A, Mansfield J, Schreiber S, Lewis CM, Mathew CG. The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. Gastroenterology 2002;122:867–874.
13. Daly MJ, Pearce AV, Farwell L, Fisher SA, Latiano A, Prescott NJ, Forbes A, Mansfield J, Sanderson J, Langelier D, Cohen A, Bitton A, Wild G, Lewis CM, Annese V, Mathew CG, Rioux JD. Association of DLG5 R30Q variant with inflammatory bowel disease. Eur J Hum Genet 2005;13:835–839.
14. Fisher SA, Hampe J, Onnie CM, Daly MJ, Curley C, Purcell S, Sanderson J, Mansfield J, Annese V, Forbes A, Lewis CM, Schreiber S, Rioux JD, Mathew CG. Direct or indirect association in a complex disease: the role of SLC22A4 and SLC22A5 functional variants in Crohn disease. Hum Mutat 2006;27:778–785.
15. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21: 263–265.
16. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. Science 2002;296:2225–2229.
17. Hunt KA, Monsuur AJ, McArdle WL, Kumar PJ, Travis SP, Walters JR, Jewell DP, Strachan DP, Playford RJ, Wijmenga C, van Heel DA. Lack of association of MYO9B genetic variants with coeliac disease in a British cohort. Gut 2006;55:969–972.
18. Post PL, Tyska MJ, O'Connell CB, Johung K, Hayward A, Mooseker MS. Myosin-IXb is a single-headed and processive motor. J Biol Chem 2002;277:11679–11683.
19. Post PL, Bokoch GM, Mooseker MS. Human myosin-IXb is a mechanochemically active motor and a GAP for rho. J Cell Sci 1998;111:941–950.
20. Todd JA. Statistical false positive or true disease pathway? Nat Genet 2006;38:731–733.
21. Neale BM, Sham PC. The future of association studies: gene-based analysis and replication. Am J Hum Genet 2004;75:353–362.
22. Freimer N, Sabatti C. The use of pedigree, sib-pair and association studies of common diseases for genetic mapping and epidemiology. Nat Genet 2004;36:1045–1051.
23. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 2005;6:109–118.
24. Nalavadi V, Nyitrai M, Bertolini C, Adamek N, Geeves MA, Bahler M. Kinetic mechanism of myosin IXB and the contributions of two class IX-specific regions. J Biol Chem 2005;280:38957–38968.
25. Saeki N, Tokuo H, Ikebe M. BIG1 is a binding partner of myosin IXb and regulates its Rho-GTPase activating protein activity. J Biol Chem 2005;280:10128–10134.
26. Matter K, Aijaz S, Tsapara A, Balda MS. Mammalian tight junctions in the regulation of epithelial differentiation and proliferation. Curr Opin Cell Biol 2005;17:453–458.
27. Graf B, Bahler M, Hilpela P, Bowe C, Adam T. Functional role for the class IX myosin myr5 in epithelial cell infection by Shigella flexneri. Cell Microbiol 2000;2:601–616.
28. Russo JM, Florian P, Shen L, Graham WV, Tretiakova MS, Gitter AH, Mrsny RJ, Turner JR. Distinct temporal-spatial roles for rho kinase and myosin light chain kinase in epithelial purse-string wound closure. Gastroenterology 2005;128:987–1001.

Received June 16, 2006. Accepted August 17, 2006.

Address requests for reprints to: Professor David van Heel, Centre for Gastroenterology, Institute of Cell and Molecular Science, Barts & The London, Queen Mary's School of Medicine & Dentistry, Turner Street, London, E1 2AD, United Kingdom. e-mail: d.vanheel@qmul.ac.uk; fax: (44) 207-882-2192.

Supported by the Netherlands Organization for Health Research and Development; the Dutch Digestive Disease Foundation; the Celiac Disease Consortium (an innovative cluster approved by the Netherlands Genomics Initiative and partly funded by the Dutch Government); the Wellcome Trust (GR068094MA); CORE; the Charity of Guy's and St

Thomas' Hospitals; and Coeliac UK. J.D.R. is supported by the National Institutes of Diabetes, Digestive and Kidney Disease and Crohn's and Colitis Federation of America. The authors acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the UK Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02.

The authors thank all individuals (patients and controls) for participating in this study. The authors also thank Dianne Soars for UK data collection.

A.A.V.B., C.R.C., K.A.H., and A.J.M. contributed equally; and C.W., M.J.D., J.D.R., and D.A.V.H. also contributed equally to this article.

Chapter 3.5

Association of tight junction genes *PARD3* and *MAGI2* with gluten-sensitive enteropathy and ulcerative colitis implies a common barrier defect

Martin C Wapenaar, Alienke J Monsuur, Ad A van Bodegraven, Rinse K Weersma, Marianna R Bevova, Ronald K Linskens, Peter Howdle, Geoffrey Holmes, Chris J Mulder, Gerard Dijkstra, David A van Heel, and Cisca Wijmenga

Submitted

Association of tight junction genes *PARD3* and *MAGI2* with gluten-sensitive enteropathy and ulcerative colitis implies a common barrier defect

Martin C Wapenaar[1,2], Alienke J Monsuur[1], Ad A van Bodegraven[3], Rinse K Weersma[4], Marianna R Bevova[1], Ronald K Linskens[5], Peter Howdle[6], Geoffrey Holmes[7], Chris J Mulder[3], Gerard Dijkstra[4], David A van Heel[8], and Cisca Wijmenga[1,2]

[1]Complex Genetics Section, DBG-Department of Medical Genetics, University Medical Center Utrecht, Utrecht, the Netherlands; [2]Genetics Department, University Medical Center Groningen, Groningen, The Netherlands. [3]Department of Gastroenterology and Hepatology, VU University Medical Center, Amsterdam, the Netherlands; [4]Department of Gastroenterology and Hepatology, University Medical Center Groningen and University of Groningen, Groningen, the Netherlands; [5]Department of Gastroenterology and Internal Medicine, St Anna Ziekenhuis, Geldrop, The Netherlands; [6]St James University Hospital, Leeds, UK; [7]Derby Royal Infirmary, Derby, UK; [8]Institute of Cell and Molecular Science, Barts and The London, Queen Mary's School of Medicine and Dentistry, London, UK. *Corresponding author:* Prof. Cisca Wijmenga, Genetics Department, room E2.030, University Medical Center Groningen, P.O. Box 30001, 9700 RB Groningen, c.wijmenga@medgen.umcg.nl, c.wijmenga@umcutrecht.nl

Abstract

Background & Aims: Celiac disease (GSE) and inflammatory bowel disease (IBD) are common gastrointestinal disorders. Both display enhanced intestinal permeability, initiated by gluten exposure (GSE) or bacterial interactions (IBD). Previous studies showed association of both diseases with variants in *MYO9B*, presumably involved in epithelial permeability. We hypothesized that genetic variants in tight junction genes might affect epithelial barrier function, thus contributing to a shared pathogenesis of GSE and IBD.

Methods: We tested this hypothesis with a comprehensive genetic association analysis of 41 genes from the tight junction pathway, represented by 201 tag SNP markers.

Results: Two genes, *PARD3* and *MAGI2*, showed association with GSE in a Dutch cohort. Replication in a British GSE cohort and joint analysis of UK and Dutch data further substantiated the association for both *PARD3* ($P = 3.2 \times 10^{-5}$; OR 1.23, 95% CI 1.11-1.37) and *MAGI2* ($P = 3.8 \times 10^{-4}$; OR 1.19, 95% CI 1.08-1.32). Association was also observed in Dutch ulcerative colitis patients with *PARD3* ($P = 0.034$; OR 1.17, 95% CI 0.99-1.38) and *MAGI2* ($P = 0.0018$; OR 1.26, 95% CI 1.08-1.47).

Conclusions. These results suggest that celiac disease and ulcerative colitis may share a common etiology through tight junction-mediated barrier defects.

Introduction

Gluten-sensitive enteropathy (GSE), or celiac disease, and inflammatory bowel disease (IBD) are two common gastrointestinal inflammatory disorders, both showing enhanced intestinal epithelial permeability (1, 2). Some healthy first-degree relatives of IBD patients also display an impaired barrier function, suggesting that this is a heritable feature rather than acquired (3). In further support of this, it has been shown that in IBD cases in human and mouse, and in GSE cases in human and dog, an impaired intestinal permeability is present long before the onset of disease (4-7). Moreover, treatment of GSE patients on a gluten-free diet reverses the disease process but does not completely reduce the increased intestinal permeability (8). Functional abnormalities of tight junctions have been observed in non-inflamed ileum of Crohn's disease patients as well, compatible with a role in IBD pathogenesis (9). Similarly, altered expression, localization, and phosphorylation of epithelial junctional proteins have been observed in GSE (10). This could mean that the barrier defect observed in GSE and IBD is mediated through epithelial tight junctions.

Both GSE and IBD have a strong genetic component and there are currently multiple susceptibility loci in the human genome linked to either IBD (reviewed in Newman et al., 2005) or GSE (reviewed in Van Heel et al., 2005) (11, 12). These two studies show sharing of certain chromosomal regions that may predispose to both disorders, e.g. a locus on 5q31-q33 and a locus on 19p13 (13). GSE and IBD co-occur in families and patients, with an approximately five-fold increased prevalence of IBD in GSE patients (14-17). This suggests that IBD and GSE share part of their genetic susceptibility.

Based on the above, genes encoding tight junction proteins should be considered highly relevant functional candidate genes for GSE and IBD. In this study we focused on 41 genes from the tight junction pathway, including those encoding for transmembrane, adaptor, signal transduction, and transcriptional-regulatory proteins (18, 19). By using genetic association analysis with a tag SNP approach we identified two adaptor protein-coding genes involved in celiac disease in Dutch and British patients. One of these genes was also associated with ulcerative colitis in a Dutch patient cohort. This suggests that both disorders share genetic risk factors that point to the involvement of the epithelial intestinal barrier.

Methods

Patients and controls

The cohorts used in this study are summarized in supplementary table 1. DNA, isolated from whole blood, was available from a cohort of 463 Dutch celiac disease patients

and 752 British celiac disease patients (20-21). The Dutch IBD case cohort has been described in detail elsewhere (22, 23). Two cohorts of random Dutch blood bank donors were available: Set 1 controls (n = 470, described by Monsuur et al., 2005)and set 2 controls (n = 459, described by van Bodegraven et al.) (20, 23). The 1185 British controls were lymphoblastoid cell line DNA from the 1958 British Birth Cohort. All Dutch cases and controls were from the Netherlands, of European descent, and with at least 3 of the 4 grandparents also born in the Netherlands. This study was approved by the Medical Ethical Committee of the University Medical Center Utrecht, the VU Medical Center Amsterdam Ethics Committee, University Medical Center Groningen Ethics Committee, and the Oxford Research Ethics Committee.

Comprehensive screen: tag SNP selection and genotyping

We selected 44 tight junction genes based on the tight junction network described in the literature (18, 19, 24), and online (KEGG: www.genome.jp/kegg/). Three genes were excluded for further study: no tag SNPs could be selected for *PARD6A* and *RAB13*, while *CLDN2* was localized on the X-chromosome. SNPs were selected by downloading all the SNPs typed in the CEPH population (Utah residents with ancestry from northern and western Europe) that were located in the genomic sequence of the remaining 41 genes from the HapMap database (November 2004, Phase I, http://www.hapmap.org/) (25). The program Tagger (available at http://www.broad.mit.edu/mpg/tagger/) was subsequently used to select tag SNPs, such that all SNPs with a minor allele frequency (MAF) $\geq$ 10% were captured with $r^2 \geq 0.7$ (excluding SNPs with low Illumina quality design scores). A few genes required too many tag SNPs to cover the entire genomic sequence. For these we tagged only the exons and exon-intron boundaries and used a MAF $\geq$ 20% (see supplementary table 2). A final set of 215 tag SNPs was obtained for genotype analysis of 463 Dutch celiac cases and set 1 controls (n=470). See supplementary table 3 for a detailed list of the tag SNPs used.

SNP genotyping was performed using the GoldenGate assay on an Illumina BeadStation 500 GX (Illumina Inc., San Diego, USA). All tag SNPs were examined for their resulting quality and 13 SNPs which had a low signal or too wide clusters were excluded. One tag SNP was not in Hardy-Weinberg equilibrium (HWE) in the controls and was therefore excluded. This yielded a total of 201 SNPs that were successfully analyzed (93%).

The five SNPs that revealed $P < 0.01$ in the comprehensive screen were genotyped as Assays on Demand (Applied Biosystems, Foster City, California, USA) in set 2 of the Dutch controls (n=459), the British celiac cases (n=752) and controls (n=1185), and the Dutch IBD cases (n=1112). Due to technical problems the second part of the IBD cohort failed for SNP rs1496770. Therefore, we only show the results for the

first part of the cohort (IBD n=588, Crohn's disease n= 298, ulcerative colitis n=290) for this particular SNP. No allele frequency differences were observed between the Dutch control cohorts (see supplementary table 4). In addition, no population stratification was observed between Dutch GSE cases and Dutch controls (De Kovel et al., unpublished observations).

Statistical analysis

Association chi-squares and two-tailed *P* values were calculated using the Haploview program (freely available at http://www.hapmap.org), for each stage of the study. SNPs that were not in HWE ($P \leq 0.001$) in the controls were excluded from further analysis. We used multiple logistic regression analysis to estimate allelic and genotypic odds ratios (OR) and the corresponding 95% confidence intervals for the five SNPs tested in the follow-up studies. The analyses were performed using STATA statistical software, version 8.0 for MS Windows. The pooled analysis of the Dutch and British GSE cohorts was performed using Mantel-Haenszel statistics.

Results

Primary genetic screen with Dutch GSE patients

In a first screen, 201 tag SNPs selected from 41 genes from the tight junction network were genotyped in a cohort of 463 Dutch GSE cases and 470 Dutch controls (see supplementary table 5). We observed evidence in favor of association ($P_{uncorrected} < 0.01$) for five SNPs in two genes (*PARD3* located on 10p11, and *MAGI2* located on 7q21) (table 1). *PARD3* and *MAGI2* both encode tight junction adaptor proteins that act as membrane-associated scaffolds. Given that these 41 tight junction genes were selected based on prior knowledge, it is not certain what level of proof one would require to proof association. When corrected for multiple testing 41 functional candidate genes, SNP rs6962966 in *MAGI2* remained significant ($P_c < 0.05/41 = 0.0012$).

Follow-up genetic analysis including British GSE patients

A follow-up study of the five SNPs with $P < 0.01$ was performed by typing 459 extra Dutch controls, and further a second, fully independent, cohort of 752 unrelated British GSE patients and 1185 British controls. The allele frequencies between the three different control cohorts (two Dutch, one British) did not differ significantly for all five SNPs tested. Association analysis of the 463 Dutch GSE cases *versus* 929 Dutch controls improved the *P*-values for four of the SNPs, two in each gene (see table 1). The three SNPs in *MAGI2* were not in linkage disequilibrium with each

Table 1. Association analysis of SNPs in *PARD3* and *MAGI2* in gluten-sensitive enteropathy.

	GSE cases			Controls			*P* by cohort*
	Allele counts			Allele counts			
	Major	Minor	MAF	Major	Minor	MAF	
PARD3 rs10763976							
Dutch	417	439	51.3	510	402	44.1	0.0024
Dutch extra controls	417	439	51.3	1014	782	43.5	0.0001
UK	717	749	51.1	1221	1105	47.5	0.016
Pooled Dutch + UK	1134	1188	51.2	2235	1887	45.8	$3.2x10^{-5}$
PARD3 rs4379776							
Dutch	538	316	37.0	636	276	30.3	0.0027
Dutch extra controls	538	316	37.0	1233	561	31.3	0.0017
UK	906	556	38.0	1496	814	35.2	0.041
Pooled Dutch + UK	1444	872	37.7	2729	1375	33.5	0.0008
MAGI2 rs6962966							
Dutch	397	459	53.6	494	418	45.8	0.0011
Dutch extra controls	397	459	53.6	960	840	46.7	0.0004
UK	744	724	49.3	1235	1077	46.6	0.050
Pooled Dutch + UK	1141	1183	50.9	2195	1917	46.6	0.00038
MAGI2 rs9640699							
Dutch	484	372	43.5	574	338	37.1	0.0061
Dutch extra controls	484	372	43.5	1115	685	38.1	0.0038
UK	915	565	38.2	1421	883	38.3	0.46
Pooled Dutch + UK	1399	937	40.1	2536	1568	38.2	0.055
MAGI2 rs1496770							
Dutch	464	392	45.8	550	362	39.7	0.0095
Dutch extra controls	464	392	45.8	1063	739	41.0	0.0099
UK	852	582	40.6	1339	923	40.8	0.4475
Pooled Dutch + UK	1316	974	42.5	2402	1662	40.9	0.085

*The initial analysis of 483 Dutch GSE cases and 470 Dutch controls were two-tailed P values calculated using Haploview. On the basis of these findings we determined *a priori* the A variant of rs10763976, the A variant of rs4379776, the G variant of rs6962966, the A variant of rs9640699, and the A variant of rs1496770 as the reference alleles for further analysis. This allowed us to perform one-sided hypothesis testing on the follow-up cohorts (i.e. the extra Dutch controls (Sets 1 and 2), the UK cohorts and the pooled data).

other (D' < 0.06, r^2 < 0.003), and only weak correlation was observed between the two SNPs in *PARD3* (D' = 0.85, r^2 = 0.40). Although the change in allele frequencies in the British GSE cases was in a similar direction as the Dutch GSE cases, it was less pronounced. The *PARD3* SNPs rs10763976 (*P* = 0.016) and rs4379776 (*P* = 0.041) showed significant association, while the *MAGI2* SNP rs6962966 (*P* = 0.050) was borderline significant. Combining the Dutch and British cohorts using a Mantel-Haenszel meta-analysis (in total 1215 cases *versus* 2114 controls) strengthened the association considerably and revealed highly significant association for three of the five SNPs (table 1). We observed the smallest *P*-value for rs10763976 ($P = 3.2 \times 10^{-5}$) located in intron 21 of *PARD3*. Individuals carrying the A allele have a modest but significantly higher risk of developing GSE (OR 1.23; 95% CI 1.11 – 1.37). The most significantly associated SNP in *MAGI2* was rs6962966 ($P = 3.8 \times 10^{-4}$) located in intron 14; this SNP was associated with a 1.2-fold increased risk for GSE (OR 1.19; 95% CI 1.08 – 1.32). Since both these associated SNPs are intronic and no direct influence on protein function was predicted, they are more likely to be markers of disease susceptibility rather than being the actual causal functional variants.

Genetic association in Dutch IBD patients

We were then interested to test if these two tight junction adaptor proteins were also associated with IBD as a group, or its clinical sub-phenotypes Crohn's disease and ulcerative colitis. In total 607 Dutch Crohn's disease and 505 Dutch ulcerative colitis cases were genotyped with the same five tag SNPs in *PARD3* and *MAGI2*. The ulcerative colitis group showed significant association to rs6962966 in *MAGI2* (*P* = 0.0018; OR 1.26, 95% CI 1.08 - 1.47) and weaker, though significant, association to rs4379776 in *PARD3* (*P* = 0.034; OR 1.17, 95% CI 0.99 - 1.38) (table 2). No association was observed for Crohn's disease.

Discussion

Weakening of the barrier function of tight junctions in the intestinal wall can result from natural interaction with pathogens (26, 27), as well as from food components such as gluten (28, 29). Our study shows for the first time that variants in tight junction genes contribute to the pathogenesis of the dissimilar gastrointestinal disorders celiac disease and ulcerative colitis. The two associated genes, *PARD3* and *MAGI2*, encode adaptor proteins that are involved as scaffolding proteins in tight junction assembly. Several of the membrane-associated proteins of the tight junction complex interact with the actin cytoskeleton. Signal transduction to the actin cytoskeleton is important in regulating both tight junction assembly and function, with a pivotal role for the

Table 2. Association analysis of SNPs in *PARD3* and *MAGI2* in inflammatory bowel disease.

IBD Combined	**IBD cases**			**Controls**			***P* by cohort***
	Allele counts			Allele counts			
	Major	Minor	MAF	Major	Minor	MAF	
PARD3 rs10763976							
IBD	1188	954	44.5	1014	782	43.5	0.27
Crohn's disease	643	525	44.9	1014	782	43.5	0.23
Ulcerative colitis	545	429	44.0	1014	782	43.5	0.40
PARD3 rs4379776							
IBD	1435	709	33.1	1233	561	31.3	0.11
Crohn's disease	796	370	31.7	1233	561	31.3	0.40
Ulcerative colitis	639	339	34.7	1233	561	31.3	0.034
MAGI2 rs6962966							
IBD	1060	1080	50.5	960	840	46.7	0.0087
Crohn's disease	595	567	48.8	960	840	46.7	0.13
Ulcerative colitis	465	513	52.5	960	840	46.7	0.0018
MAGI2 rs9640699							
IBD	1307	843	39.2	1115	685	38.1	0.23
Crohn's disease	722	448	38.3	1115	685	38.1	0.45
Ulcerative colitis	585	395	40.3	1115	685	38.1	0.12
MAGI2 rs1496770							
IBD	670	452	40.3	1063	739	41.0	0.35
Crohn's disease	339	231	40.5	1063	739	41.0	0.42
Ulcerative colitis	331	221	40.0	1063	739	41.0	0.34

*The initial analysis in GSE was based on a two-sided analysis using Haploview. On the basis of these findings we determined *a priori* the A variant of rs10763976, the A variant of rs4379776, the G variant of rs6962966, the A variant of rs9640699, and the A variant of rs1496770 as the reference alleles for further analysis. This allowed us to perform a one-sided hypothesis testing on the follow-up cohorts. Due to technical problems, the second part of the IBD cohort failed for rs1496770, we can therefore only show the results of the first part of the cohort (IBD n=588, Crohn's disease n= 298, ulcerative colitis n=290).

small GTPase RhoA (30, 31). Myosin IXB might also play a role in this process, as it possesses a Rho-GAP domain that negatively regulates Rho proteins. Recently, we demonstrated that the GSE-associated gene *MYO9B* (20), which localizes to the *IBD6* locus on chromosome 19p13, was associated with ulcerative colitis (UC) and to a lesser extent to Crohn's disease (which together comprise the IBD phenotype) in four different Caucasian populations, i.e. Dutch, British, Italian, and French-Canadian (23). We hypothesized that myosin IXB – which contains a Rho-GTPase activating domain (GAP) – is involved in intestinal permeability through remodeling of the epithelial cytoskeleton and tight junction assembly, through its interaction with RhoA (20, 23). Since the primary barrier defects are expected to be subtle, we presume that the cycle of barrier malfunction is sustained by the subsequent local inflammations, since it is well established that proinflammatory cytokines – such as interferon-gamma and tumor necrosis factor – trigger further barrier dysfunction.

Both GSE and ulcerative colitis are characterized by a superficial inflammation restricted to the intestinal mucosa. In this respect it is interesting to note that the epithelial tight junction-related genes *MYO9B, PARD3* and *MAGI2* were collectively associated with both GSE and ulcerative colitis. A healthy gut mucosa is characterized by a robust and selective barrier, maintained by properly differentiated and polarized epithelial cells. Weakening of the barrier due to external factors like gluten or intestinal bacteria in conjunction with enhanced sensitivity of the host, not only results in inflammation but also leads to a loss of epithelial differentiation and polarization, resulting in clinical phenotypes characterized by nutrient malabsorption and diarrhea. The outcome of the interaction between potentially harmful external factors and the mucosal barrier may be influenced by genetic variants in genes that code for the tight junction complex. In support of our finding is the recently reported association between IBD and the gene *DLG5,* which encodes a scaffold protein involved in maintenance of epithelial cell contacts and polarity (32). Its impairment is also considered to interfere with the gut barrier function.

In conclusion, we have demonstrated that the genes *PARD3* and *MAGI2* (together with the previously reported *MYO9B*) are genetically associated with both celiac disease and ulcerative colitis, suggesting they share a common etiology through tight junction-mediated intestinal barrier impairment.

Acknowledgements

We thank all the patients, their physicians, and the Dutch Celiac Disease Foundation for participating in this study. We thank Dalila Pinto for genotyping part of the IBD cohort, Harry van Someren as database manager, and Jackie Senior for critically

reading the manuscript. We would like to thank Karen A Hunt (Barts and the London) for extracting the DNA and providing all the samples from the British cohort. We acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the UK Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02, and the help of Prof David Strachan (St Georges University of London) in acquiring these samples. We also thank Prof Raymond J Playford and Prof Parveen Kumar (Barts and the London), Prof Derek P Jewell, Dr Simon PL Travis (John Radcliffe Hospital, Oxford), Dr Julian RF Walters (Hammersmith Hospital, London), Dr Jill Swift and Dr Rosemary Crimmins (Llandough Hospital, Cardiff). This study was supported by grants from the Netherlands Organization for Scientific Research (grants 902-22-094 and 912-02-028), the Dutch Digestive Diseases Foundation (grants 97-44 and 03-06) and the Celiac Disease Consortium, an innovative cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch government (grant BSIK03009), and the Wellcome Trust (Clinician Scientist Fellowship GR068094MA) and Coeliac UK.

References

1. Fasano A, Shea-Donohue T. Mechanisms of disease: the role of intestinal barrier function in the pathogenesis of gastrointestinal autoimmune diseases. Nat Clin Pract Gastroenterol Hepatol 2005;2:416-22.
2. Schreiber S, Rosenstiel P, Albrecht M, et al. Genetics of Crohn disease, an archetypal inflammatory barrier disease. Nat Rev Genet 2005;6:376-88.
3. Halme L, Paavola-Sakki P, Turunen U, et al. Family and twin studies in inflammatory bowel disease. World J Gastroenterol 2006;12:3668-72.
4. Ma TY. Intestinal epithelial barrier dysfunction in Crohn's disease. Proc Soc Exp Biol Med 1997;214:318-27.
5. Sugawara K, Olson TS, Moskaluk CA, et al. Linkage to peroxisome proliferator-activated receptor-gamma in SAMP1/YitFc mice and in human Crohn's disease. Gastroenterology 2005;128:351-60.
6. Schulzke JD, Schulzke I, Fromm M, et al. Epithelial barrier and ion transport in coeliac sprue: electrical measurements on intestinal aspiration biopsy specimens. Gut 1995;37:777-82.
7. Hall EJ, Batt RM. Abnormal permeability precedes the development of a gluten sensitive enteropathy in Irish setter dogs. Gut 1991;32:749-53.
8. Smecuol E, Sugai E, Niveloni S, et al. Permeability, zonulin production, and enteropathy in dermatitis herpetiformis. Clin Gastroenterol Hepatol 2005;3:335-41.
9. Soderholm JD, Olaison G, Peterson KH, et al. Augmented increase in tight junction permeability by luminal stimuli in the non-inflamed ileum of Crohn's disease. Gut 2002;50:307-13.
10. Ciccocioppo R, Finamore A, Ara C, et al. Altered expression, localization, and phosphorylation of epithelial junctional proteins in celiac disease. Am J Clin Pathol 2006;125:502-11.
11. Newman B, Siminovitch KA. Recent advances in the genetics of inflammatory bowel disease. Curr Opin Gastroenterol 2005;21:401-7.

12. van Heel DA, Hunt K, Greco L, et al. Genetics in coeliac disease. Best Pract Res Clin Gastroenterol 2005;19:323-39.
13. Curley CR, Monsuur AJ, Wapenaar MC, et al. A functional candidate screen for coeliac disease genes. 2006.
14. Cottone M, Marrone C, Casa A, et al. Familial occurrence of inflammatory bowel disease in celiac disease. Inflamm Bowel Dis 2003;9:321-3.
15. Schedel J, Rockmann F, Bongartz T, et al. Association of Crohn's disease and latent celiac disease: a case report and review of the literature. Int J Colorectal Dis 2005;20:376-80.
16. Tursi A, Giorgetti GM, Brandimarte G, et al. High prevalence of celiac disease among patients affected by Crohn's disease. Inflamm Bowel Dis 2005;11:662-6.
17. Yang A, Chen Y, Scherl E, et al. Inflammatory bowel disease in patients with celiac disease. Inflamm Bowel Dis 2005;11:528-32.
18. Matter K, Balda MS. Signalling to and from tight junctions. Nat Rev Genet 2003;4:225-36.
19. Schneeberger EE, Lynch RD. The tight junction: a multifunctional complex. Am J Physiol Gastrointest Liver Physiol 2004;286:C1213-28.
20. Monsuur AJ, de Bakker PI, Alizadeh BZ, et al. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. Nature genetics 2005;37:1341-4.
21. Hunt KA, Monsuur AJ, McArdle W, et al. Lack of association of MYO9B genetic variants with coeliac disease in a British cohort. 2006.
22. Oostenbrug LE, Drenth JP, de Jong DJ, et al. Association between Toll-like receptor 4 and inflammatory bowel disease. Inflamm Bowel Dis 2005;11:567-75.
23. Van Bodegraven AA, Curley CR, Hunt K, et al. Genetic variation in myosin IXB is associated with ulcerative colitis. Gastroenterology 2006;in press.
24. Matter K, Aijaz S, Tsapara A, et al. Mammalian tight junctions in the regulation of epithelial differentiation and proliferation. Curr Opin Cell Biol 2005;17:453-8.
25. The International HapMap Project. Nature 2003;426:789-96.
26. Berkes J, Viswanathan VK, Savkovic SD, et al. Intestinal epithelial responses to enteric pathogens: effects on the tight junction barrier, ion transport, and inflammation. Gut 2003;52:439-51.
27. Clayburgh DR, Shen L, Turner JR. A porous defense: the leaky epithelial barrier in intestinal disease. Lab Invest 2004;84:282-91.
28. Drago S, El Asmar R, Di Pierro M, et al. Gliadin, zonulin and gut permeability: Effects on celiac and non-celiac intestinal mucosa and intestinal cell lines. Scand J Gastroenterol 2006;41:408-19.
29. Sander GR, Cummins AG, Henshall T, et al. Rapid disruption of intestinal barrier function by gliadin involves altered expression of apical junctional proteins. FEBS Lett 2005;579:4851-5.
30. Rajasekaran SA, Palmer LG, Moon SY, et al. Na,K-ATPase activity is required for formation of tight junctions, desmosomes, and induction of polarity in epithelial cells. Mol Biol Cell 2001;12:3717-32.
31. Takakuwa R, Kokai Y, Kojima T, et al. Uncoupling of gate and fence functions of MDCK cells by the actin-depolymerizing reagent mycalolide B. Exp Cell Res 2000;257:238-44.
32. Stoll M, Corneliussen B, Costello CM, et al. Genetic variation in DLG5 is associated with inflammatory bowel disease. Nature genetics 2004;36:476-80.

Part 4

Inflammation in CD and screening for HLA

Chapter 4.1

A functional candidate screen for coeliac disease genes

Alienke J Monsuur, Christine R Curley, Martin C Wapenaar, John D Rioux and Cisca Wijmenga

European Journal of Human Genetics 2006; 14:1215–1222

ARTICLE

A functional candidate screen for coeliac disease genes

Christine R Curley[1,4], Alienke J Monsuur[2,4], Martin C Wapenaar[2], John D Rioux[1,3] and Cisca Wijmenga*[,2]

[1]*The Broad Institute, Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA;* [2]*Complex Genetics Section, DBG-Department of Medical Genetics, University Medical Centre Utrecht, Utrecht, The Netherlands;* [3]*Université de Montreal and the Montreal Heart Institute/Institut de Cardiologie de Montreal, Montreal, Quebec, Canada*

It is increasingly evident that different inflammatory disorders show some overlap in their pathological features, concurrence in families and individuals, and shared genetic factors. This might also be true for coeliac disease, a chronic inflammatory disorder of the gastrointestinal system, which shares two linkage regions with inflammatory bowel disease: on chromosome 5q31 (CELIAC2 and IBD5) and 19p13 (CELIAC4 and IBD6). We hypothesised that these regions contain genes that contribute to susceptibility to both disorders. The overlapping 5q31 region contains only five positional candidate genes, whereas the overlapping 19p13 region has 141 genes. As the common disease gene probably plays a role in inflammation, we selected five functional candidate genes from the 19p13 region. We studied these 10 positional and functional candidate genes in our Dutch coeliac disease cohort using 44 haplotype tagging single-nucleotide polymorphisms. Two genes from 19p13 showed a small effect on familial clustering: the cytochrome *P*450 F3 gene *CYP4F3* ($P_{nominal}$ 0.0375, odds ratio (OR) 1.77) and *CYP4F2* ($P_{nominal}$ 0.013, OR 1.33). CYP4F3 and CYP4F2 catalyse the inactivation of leukotriene B4 (LTB4), a potent mediator of inflammation responsible for recruitment and activation of neutrophils. The genetic association of LTB4-regulating gene variants connects the innate immune response of neutrophil mobilisation with that of the established Th1 adaptive immunity present in coeliac disease patients. The findings in coeliac disease need to be replicated. Expanding genetic association studies of these cytochrome genes to other inflammatory conditions should reveal whether their causative influence extends beyond coeliac disease.
European Journal of Human Genetics (2006) **14,** 1215–1222. doi:10.1038/sj.ejhg.5201687; published online 12 July 2006

Keywords: coeliac disease; inflammatory bowel disease; *MYO9B*; inflammation; candidate gene study; IBD5

Introduction

Coeliac disease is a chronic disease characterised by an inflammatory response in the gastrointestinal tract and an impaired intestinal epithelial barrier.[1] It is a complex genetic disorder, involving multiple genetic variants. Genetic studies on coeliac disease patients indicate that multiple chromosomal regions predispose to disease susceptibility, confirming the suggestion that common genes contribute to this disorder. Two of the coeliac disease loci, CELIAC2 on chromosome 5q23–q33 and CELIAC4 on 19p13.1,[2,3] coincide with linkage regions for inflammatory bowel disease (IBD),[4,5] giving rise to the hypothesis of common disease susceptibility. The IBD5 and IBD6 loci on 5q31 and 19p13.1, respectively, are among the most significant and consistently

*Correspondence: Professor C Wijmenga, Complex Genetics Section, Stratenum 2.117, DBG-Department of Biomedical Genetics, University Medical Centre Utrecht, PO Box 85060, 3508 AB Utrecht, The Netherlands. Tel: +31 30 253 8427; Fax: +31 30 253 8479;
E-mail: t.n.wijmenga@med.uu.nl
[4]These authors have contributed equally to this work.
Received 14 December 2005; revised 18 May 2006; accepted 30 May 2006; published online 12 July 2006

replicated IBD loci.[6,7] There is ample evidence that (auto)immune-related disorders share some of their genetic susceptibility factors, the best examples being variations in the *CTLA-4* gene[8] and *PTPN22*,[9–11] both of which help regulate T-cell responsiveness.

IBD, clinically classified as Crohn's disease and ulcerative colitis, also shows a chronic inflammation of the gastrointestinal tract.[6] Coeliac disease and IBD co-occur in families and patients, with an approximately five-fold increased prevalence of IBD in coeliac disease patients,[12–15] further suggesting common pathophysiological mechanisms for both diseases.

The sharing of the linkage regions, the concurrence in patients and the pathological commonalities between coeliac disease and IBD led us to hypothesise that genes on chromosomes 5q31 and 19p13.1 might be associated with susceptibility for both disorders, most probably by influencing inflammation in the gut. The CELIAC2 locus on 5q23–q33 overlaps with a 250 kb haplotype on 5q31 associated with Crohn's disease (IBD5).[16] The implicated haplotype contains five genes (*P4HA2*, *IRF1*, *SLC22A5/OCTN2*, *SLC22A4/OCTN1*, *PDLIM3*) (see Table 1), all of which are positional candidate genes for coeliac disease pathogenesis. The CELIAC4 locus ranging from 15.38 to 21.08 Mb (99% confidence interval (CI), Ensembl version 35) lies completely within the IBD6 locus, which spans a large part of chromosome 19p. Of the 141 genes in the overlapping region, five are known to be involved in inflammation and were considered as functional candidate genes: *CYP4F3*, *CYP4F2*, *HSH2D*, *IL12RB1* and *IFI30* (see Table 1). A complementary approach to identifying the susceptibility gene for coeliac disease in the chromosome 19 region involves fine-mapping of the region. We simultaneously undertook such a strategy, which led to

Table 1 Overview of the 10 positional and functional candidate genes studied in relation to the overlapping regions in coeliac disease and IBD

Gene name	*Chromosome*	*Location (bps)*	*Function*	*References*	*No. of tag SNPs*[a]
P4HA2	5	131 556 202–131 590 458	Prolyl 4-hydroxylase α-2 subunit precursor. Catalyses the post-translational formation of 4-hydroxyproline in –Xaa–Pro–Gly– sequences in collagens and other proteins	25	—
PDLIM4	5	131 621 285–131 637 046	PDZ and LIM domain 4. Interacts with the LIM domain with the second and fourth PDZ domains of PTPN13	16	—
SLC22A4	5	131 658 043–131 707 796	Organic cation/carnitine transporter 1 (solute carrier family 22, member 4). Sodium-ion-dependent, low-affinity carnitine transporter. Defects in SLC22A4 may be a cause of susceptibility to Crohn's disease and rheumatoid arthritis	16	5
SLC22A5	5	131 733 343–131 759 202	Organic cation/carnitine transporter 2 (solute carrier family 22). Sodium-ion-dependent, high-affinity carnitine transporter. Involved in the active cellular uptake of carnitine. Transports one sodium ion with one molecule of carnitine Defects are the cause of systemic primary carnitine deficiency and may be a cause of susceptibility to Crohn's disease	16	2
IRF1	5	131 846 678–131 854 333	Interferon regulatory factor 1. Specifically binds to the upstream regulatory region of type I IFN and IFN-inducible MHC class I genes (the interferon consensus sequence) and activates those genes. Deletion or rearrangement of IRF1 are a cause of preleukaemic myelodysplastic syndrome and of acute myelogenous leukaemia	16	—
CYP4F3	19	15 613 226–15 631 234	Member of *P*450 family of cytochromes. Catalyses the inactivation of leukotriene B4 (LTB4), a potent mediator of inflammation	26–28	8
CYP4F2	19	15 849 834–15 869 885	Member of *P*450 family of cytochromes. Catalyses the inactivation of LTB4, a potent mediator of inflammation	24,29	8
HSH2D	19	16 115 501–16 130 375	Target of the two most important signalling pathways of T-cell activation; T-cell receptor antigen signalling and costimulation of the naive T cell	30,31	5
IL12RB1	19	18 030 806–18 104 521	IL-12 receptor β chain 1, which is involved in inducing cell-mediated immunity to intracellular pathogens via the Th1 pathway and interferon production. Deficiency leads to impaired mycobacterial immunity. Large amounts of IL-12 are secreted in mice with an NOD2 mutation in response to TLR2, which facilitates a Th1 response	32–34	3
IFI30	19	18 145 579–18 149 927	Induced by interferon-γ	35	2

[a]Seven of the 16 haploblock tagging SNPs on chromosome 5 are located in the genes, and the others are located outside the genes.

the identification of the myosin IXB gene (*MYO9B*) as a susceptibility gene for coeliac disease.[17] The odds ratio (OR) for this gene is lower than the expected OR for the CELIAC4 locus, leaving room for another susceptibility gene in this locus. In this study, we report our association findings of the five positional candidate genes comprising the 250 kb haplotype on 5q31 and the five functional candidate genes from 19p13.1; all were tested using a haplotype block tag approach for association with coeliac disease in a cohort of Dutch patients.

Materials and methods

Samples

A case–control cohort consisting of 309 independent coeliac disease patients and 358 independent controls, all of Dutch Caucasian origin, was used for the genetic association study. Patients were diagnosed according to the ESPGHAN criteria and approximately 93% of the patients were HLA-DQ2 positive. We included for this study only coeliac disease patients with a biopsy-proven Marsh III lesion, as described by Van Belzen *et al.*[3] Blood samples were collected and DNA was isolated according to the standard laboratory procedures.[3] The study was approved by the Medical Ethics Committee of the University Medical Centre in Utrecht and informed consent was obtained from all individuals.

Candidate gene and single-nucleotide polymorphism (SNP) selection

A total of 16 haplotype tagging SNPs were selected from the 5q31 region, as described before.[16] We observed comparable haplotypes to the ones published by Rioux *et al.*[16] We also selected five functional candidate genes located in the overlapping CELIAC4 and IBD6 locus on chromosome 19 for this study, based on their known function, expression studies and HapMap SNP coverage. We selected SNPs based on block-tagging ability, validation status, allele frequency and SNP density. Previous performance rating (genotyping >90%) from the IBD cohort typing, on the Sequenom platform, was also taken into account. Haplotype tagging SNPs were selected on the basis of HapMap data as of September 2004 (www.hapmap.org) using Haploview to determine the haplotype tagging SNPs. These SNPs were selected at an early stage of the HapMap project when there was less information than available today. If tag SNPs could not be determined from HapMap data, Supplemental data sets (Programs for Genomic Applications (PGA) and previous haplotype analyses by our group (JD Rioux, personal communication)) were used for haplotype tagging SNP selection using Haploview as described above. Supplementary Table 1 lists the selected tagging SNPs per gene that were tested on the coeliac disease cohort.

Genotyping

Genomic DNA extracted from whole-blood samples was used. Genotyping assays were designed using the Sequenom Assay Design program and genotypes were obtained using the Sequenom Mass Array system at the Broad Institute of MIT and Harvard as described by Gabriel *et al.*[18] Genotyping data was analysed for Hardy–Weinberg equilibrium and allele frequency. The quality control criteria used to determine if genotyping results were successful were: a minimum of 75% genotyping success for each SNP, Hardy–Weinberg equilibrium values >0.01 and observed heterozygosity >0.5%. Of 44 SNPs tested in the region, 42 SNPs passed the above criteria (one was monomorphic and the other failed owing to Hardy–Weinberg errors).

Statistical analysis

Allele and haplotype counts in cases *versus* controls were analysed for association. A single-marker and multimarker association study for each gene in the case–control cohort was conducted by a standard χ^2 test (2×2 contingency table). Haplotypes were constructed using Haploview for the unrelated cases and for the control cohorts separately.[19] OR represents maximum-likelihood estimate of odds ratio, and the corresponding 95% CI was approximated using Woolf's method.

The relative risk for coeliac disease associated with the CELIAC4 locus base on the linkage data is calculated to be 2.3.[3] The OR associated with *MYO9B* heterozygosity is estimated to be 1.6 and homozygosity 2.3.[17] We calculated the power for this study based on a relative risk of 1.7, which gives us more than 75% power to detect a disease variant with minor allele frequencies ranging from 0.1 to 0.3, assuming a dominant inheritance, a disease prevalence of 0.1, a D' of 1 and an equal frequency of the tested SNP and the high-risk variant.

Results

Coeliac disease and IBD share linkage regions on chromosome 5q23–q33 (CELIAC2 and IBD5) and 19p13 (CELIAC4 and IBD6). We performed association studies on 10 positional and functional candidate genes to search for association with genes that might play a primary role in both disorders. Forty-four SNPs were selected to tag the haplotype blocks, thereby excluding redundant typing, and were genotyped in a cohort of 309 independent Dutch coeliac disease cases and 358 Dutch controls. Two tag SNPs failed to pass our quality control standards.

In the chromosome 5 region, one SNP showed association (P_{nominal} <0.05): rs7705826 located in SLC22A5 (P_{nominal} 0.033, OR 1.39, 95% CI 1.03–1.88). Owing to the high linkage disequilibrium (LD) in the region, we would have expected to find more than one SNP from this region to be associated. Therefore, this observation is

probably false positive also because this SNP deviates from Hardy–Weinberg equilibrium in the control population.

In the chromosome 19 region, four SNPs showed single SNP association, two of which were located in *CYP4F2* and two in *CYP4F3* (Table 2, Figure 1). These genes are located head-to-head 218 kb apart. SNPs rs7252046 ($P_{nominal}$ 0.0427, OR 1.28, 95% CI 1.01–1.61) and rs3093156 ($P_{nominal}$ 0.013, OR 1.33, 95% CI 1.06–1.66) showed association in *CYP4F2*, and SNPs rs1290622 ($P_{nominal}$ 0.0447, OR 1.58, 95% CI 1.01–2.48) and rs1290625 ($P_{nominal}$ 0.0375, OR 1.77, 95% CI 1.03–3.05) in *CYP4F3*. Heterozygotes for the most associated SNP (rs3093156 in *CYP4F2*) have a 1.6 times higher risk of coeliac disease ($P_{nominal}$ 0.021, 95% CI 1.07–2.4), whereas in homozygotes the risk increases to 1.84 ($P_{nominal}$ 0.01, 95% CI 1.15–2.93). The results obtained for these genes are independent as the r^2 between the SNPs in the genes is <0.06. Haplotype analyses of each of the two SNPs in the *CYP4F2* and *CYP4F3* genes are shown in Table 3. The other seven genes (*IRF1*, *SLC22A4/OCTN1*, *PDLIM3*, *P4HA2*, *IF130*, *IL12RB1* and *HSH2D*) showed no association.

Discussion

In the last decade, much research has been devoted to elucidating the genetic basis of complex traits such as coeliac disease. Although increasingly successful, this work has been complicated by the fact that multiple genes can be associated with a trait, but each has insufficient impact to account for the total genetic susceptibility of the disease under study. Given the extent of linkage regions and the large number of genes they usually encompass, it can be a daunting task to select the most plausible candidate gene, particularly when little or no knowledge is available on gene function or the biological process perturbed. Both the coeliac disease loci on chromosome 5q23–q33 and 19p13.1 coincide with linkage regions for IBD. We therefore hypothesised that genes related to biological processes common to both disorders would make excellent functional candidates.

In this study, we observed nominal association to a single SNP located in SLC22A5 on chromosome 5. Although it is not clear which of the five genes in this region really confer susceptibility to Crohn's disease,[16] changes in the SLC22A4 and SLC22A5 have been observed in relation to Crohn's disease,[20] which potentially have a functional relevance. The effect of the association in coeliac disease is quite modest compared to the original findings in Crohn's disease. Furthermore, the other SNPs that we tested in the region show no association despite being in near-complete LD. Given these results, we feel that the most parsimonious explanation is that this association of SLC22A5 to celiac disease is a false positive, but this will require confirmation by independent groups.

Recently, a study has been published investigating 56 candidate genes from chromosome 19 for their involvement in IBD.[21] All but one (*CYP4F2*) of the genes included in this study overlap with the genes published by Tello-Ruiz *et al*.[21] We observed association to the cytochrome genes, *CYP4F2* and *CYP4F3*, in a cohort of Dutch coeliac disease patients. Although the genes are juxtaposed on chromosome 19, the observed association signals are independent as there is hardly any LD between them (maximum r^2 between the associated SNPs of the two genes is <0.06). The relative risk associated to each of the two genes ranges from 1.6 to 1.8, which is insufficient to fully explain the linkage result observed earlier.[3] It is interesting that variants in the myosin IXB gene (*MYO9B*), located approximately 1.2 Mb proximal to *CYP4F2*, have recently been shown to be associated with increased risk (OR = 1.7) to coeliac disease.[17] The association found in the *CYP4F2* and *CYP4F3* genes is not owing to long-range LD between these genes and *MYO9B* (maximum r^2 <0.0044). Although the observation with the *CYP4F2* and *CYP4F3* genes has not been corrected for multiple testing and needs to be independently replicated, it is tempting to speculate that the original strong linkage signal on 19p13.1[3] resulted from the presence of multiple susceptibility genes.

If we look at their function, the involvement of these two *CYP4F* genes is intriguing. Both CYP4F isoforms are involved in the oxidative degradation of leukotriene B4 (LTB4), the arachidonic acid-derived lipid inflammatory mediator responsible for the recruitment and activation of neutrophils. The genetic association of LTB4-regulating gene variants connects the innate immune response of neutrophil mobilisation with that of the established Th1 adaptive immunity present in coeliac disease. These genetic variants may influence neutrophil migration and thus create an environment in the mucosa that contributes to coeliac disease pathogenesis. It has been reported that activated neutrophils have an effect on epithelial tight junctions that results in enhanced permeability[22] in the intestine, which could facilitate the influx of various commensals from the lumen leading to the recruitment of even more phagocytic neutrophils. This self-sustaining cycle of barrier impairment could also enable gluten to enter the lamina propria, where it could be presented to resident $CD4^+$ T cells to evoke the Th1 response. In turn, activation of the Th1 immune pathway would further undermine barrier integrity through the release of interferon-γ.[23] In a parallel gene expression study, we observed increased neutrophil numbers not only in untreated coeliac patients but also in patients in complete remission, probably reflecting a genetic impairment of the epithelial barrier.[25] This observation is in line with the recently identified *MYO9B* gene, which is expected to confer genetic susceptibility to coeliac disease pathogenesis by also affecting the intestinal barrier.[17] A more detailed assessment of the contribution of *CYP4F2* and *CYP4F3* variants

Table 2 Association results for the SNPs that were polymorphic and in Hardy–Weinberg equilibrium (HWE)

							Frequency[a]		*Allele counts*[b]				
Gene	*rs number*	*Chromosome*	*Base pair position*	*MAJ allele*[a]	*HWE P-value controls*	*HWE P-value cases*	*Case*	*Control*	*Case*	*Control*	χ^2	*P-value*	*OR (95% CI)*[c]
SLC22A4	rs7705189	5	131 651 257	T, T	0.227	0.802	0.552	0.532	328:266	367:323	0.530	0.467	
SLC22A4	rs162887	5	131 657 064	C, C	0.417	0.737	0.690	0.686	422:190	479:219	0.016	0.898	
SLC22A4	IGR3081ms1	5	131 658 751	T, T	0.043	0.764	0.572	0.537	348:260	382:330	1.705	0.192	
SLC22A4	IGR3066ms1	5	131 666 223	A, A	0.217	0.436	0.528	0.520	319:285	366:338	0.089	0.766	
SLC22A4	rs27060	5	131 679 710	A, A	0.392	1.000	0.606	0.615	367:239	429:269	0.111	0.740	
—	rs4705938	5	131 721 976	A, A	0.176	0.638	0.545	0.520	325:271	361:333	0.813	0.367	
—	rs671473	5	131 736 434	G, G	0.449	0.523	0.685	0.686	419:193	480:220	0.002	0.967	
—	rs274567	5	131 742 308	G, G	1.000	0.340	0.562	0.543	326:254	357:301	0.475	0.491	
SLC22A5	rs17622208	5	131 744 949	C, C	0.228	0.469	0.542	0.514	323:273	356:336	0.971	0.324	
SLC22A5	rs7705826	5	131 760 355	C, C	0.022	0.739	0.822	0.865	498:108	595:93	4.550	0.033	1.39 (1.03–1.88)
—	rs11739135	5	131 761 296	C, C	0.272	0.992	0.610	0.567	376:240	397:303	2.528	0.112	
—	rs13179841	5	131 770 719	G, G	0.504	0.502	0.934	0.938	564:40	655:43	0.116	0.734	
—	rs4475253	5	131 804 405	T, T	0.867	0.588	0.698	0.706	420:182	490:204	0.108	0.742	
—	rs4705950	5	131 821 185	G, G	0.095	1.000	0.569	0.552	340:258	384:312	0.370	0.543	
—	rs6894249	5	131 825 446	T, T	0.303	1.000	0.613	0.649	373:235	454:246	1.723	0.189	
—	rs2248116	5	131 832 246	T, T	0.225	0.896	0.577	0.554	352:258	391:315	0.718	0.397	
CYP4F3	rs4807964	19	15 609 827	G, G	1.000	0.231	0.841	0.811	461:87	555:129	1.873	0.171	
CYP4F3	rs1290617	19	15 612 897	T, T	0.233	0.789	0.591	0.623	357:247	441:267	1.385	0.239	
CYP4F3	rs1290618	19	15 613 137	C, C	0.379	0.075	0.770	0.758	442:132	532:170	0.260	0.610	
CYP4F3	rs1290622	19	15 614 423	C, C	0.838	0.860	0.947	0.918	551:31	652:58	4.029	0.045	1.58 (1.01–2.48)
CYP4F3	rs1290625	19	15 618 156	G, G	0.650	0.122	0.940	0.965	521:33	643:23	4.327	0.038	1.77 (1.03–3.05)
CYP4F3	rs1290626	19	15 618 927	C, C	1.000	0.912	0.587	0.577	338:238	411:301	0.120	0.730	
CYP4F3	rs1543284	19	15 637 785	G, G	0.906	1.000	0.575	0.572	354:262	406:304	0.011	0.917	
CYP4F3	rs1543286	19	15 637 945	C, C	0.987	0.976	0.503	0.524	289:285	373:339	0.529	0.467	
CYP4F2	rs2189784	19	15 820 200	G, G	0.187	0.834	0.610	0.613	376:240	429:271	0.008	0.927	
CYP4F2	rs2079288	19	15 825 203	T, T	0.918	0.970	0.755	0.762	447:145	538:168	0.086	0.770	
CYP4F2	rs7252046	19	15 832 473	T, T	1.000	0.217	0.708	0.655	415:171	468:246	4.107	0.043	1.28 (1.01–1.61)
CYP4F2	rs12610189	19	15 839 641	T, T	0.669	0.568	0.635	0.664	352:202	426:216	1.038	0.308	
CYP4F2	rs2108622	19	15 851 431	C, C	0.054	0.660	0.712	0.732	430:174	521:191	0.641	0.424	
CYP4F2	rs3093156	19	15 861 609	T, A	0.850	0.124	0.544	0.527	309:259	355:319	6.167	0.013	1.33 (1.06–1.66)
CYP4F2	rs3093135	19	15 865 371	T, T	0.452	0.326	0.844	0.880	449:83	523:71	3.165	0.075	
CYP4F2	rs3761014	19	15 872 763	C, C	0.382	0.401	0.832	0.844	506:102	604:112	0.312	0.577	
HSH2	rs2032882	19	16 105 817	A, A	0.151	0.639	0.910	0.879	453:45	564:78	2.824	0.093	
HSH2	rs444053	19	16 113 072	C, C	0.531	1.000	0.719	0.725	433:169	516:196	0.048	0.826	
HSH2	rs285290	19	16 120 460	C, C	0.535	0.541	0.832	0.848	476:96	599:107	0.626	0.429	
HSH2	rs2258476	19	16 126 017	A, A	0.403	0.599	0.833	0.829	498:100	585:121	0.040	0.842	
HSH2	rs681059	19	16 130 507	A, A	1.000	0.686	0.717	0.721	433:171	516:200	0.023	0.879	
IL12RB1	rs404733	19	18 030 997	A, A	0.585	0.850	0.513	0.513	311:295	363:345	0.000	0.986	
IL12RB1	rs375947	19	18 041 451	A, A	0.675	0.454	0.686	0.674	424:194	473:229	0.228	0.633	
IL12RB1	rs436857	19	18 058 635	G, G	0.476	0.525	0.798	0.803	487:123	565:139	0.036	0.849	
IFI30	rs273266	19	18 144 501	T, T	0.932	0.213	0.787	0.772	455:123	545:161	0.429	0.513	
IFI30	rs7125	19	18 149 069	A, G	0.670	0.160	0.502	0.531	306:304	360:318	1.367	0.242	

[a]Major allele in cases and controls, which may be the opposite allele.
[b]Actual allele counts for cases and controls; the major allele is shown first.
[c]Odds ratio (OR) calculated using the not-associated allele as reference.

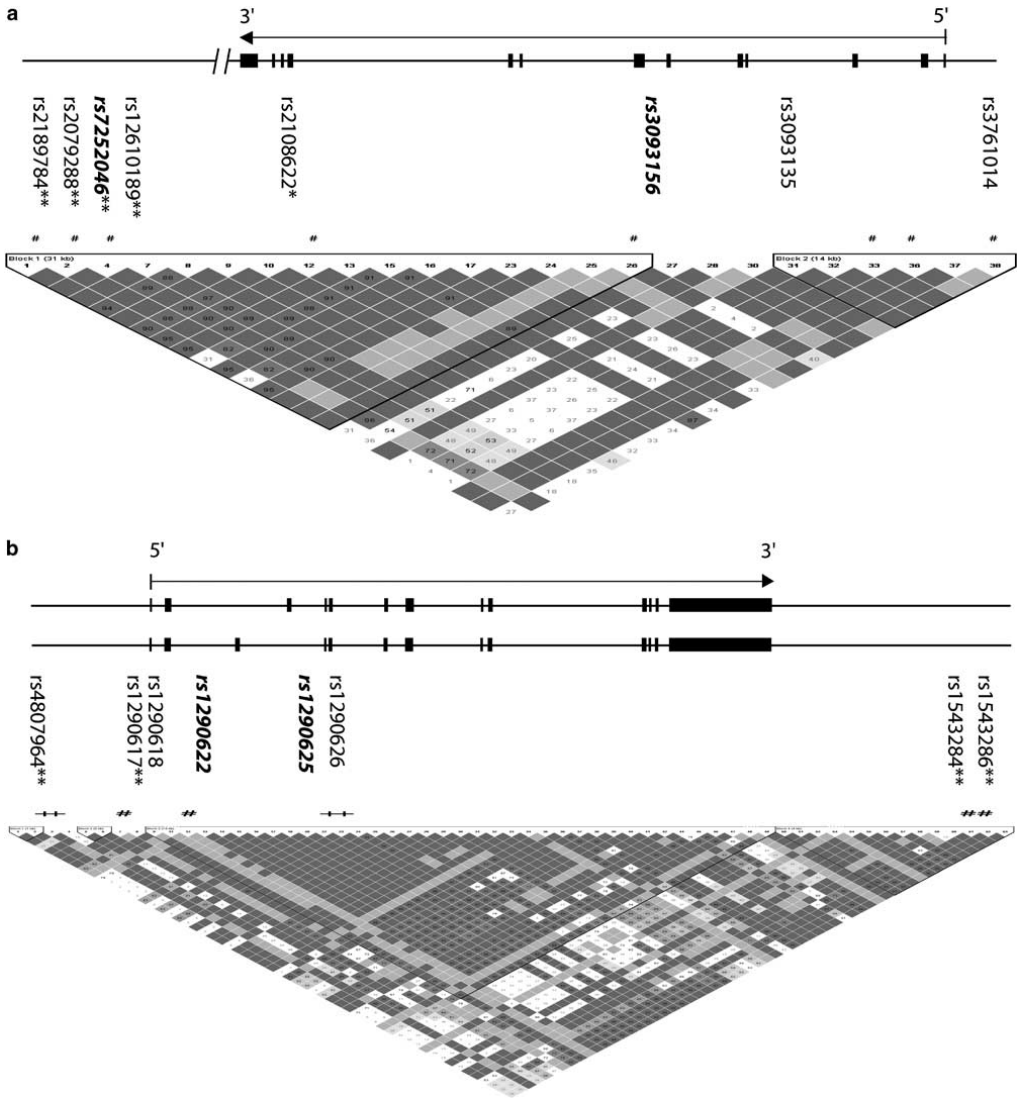


Figure 1 Overview of *CYP4F2* (**a**) and *CYP4F3* (**b**), isoforms a and b, respectively, showing the exons, the locations of the tag SNPs (associated tag SNPs are listed in bold italics) and the observed LD based on the HapMap information. **The locations of the SNPs outside the genes are not true to scale. *rs2108622 is a coding SNP; # shows the position of the SNPs in the LD plot; and + shows for the SNPs that were not typed in the HapMap, the relative position of the SNPs in the LD plot.

to neutrophil recruitment and intestinal permeability will require further functional analysis. Replication in coeliac disease and expanding genetic association studies of these cytochrome genes to other inflammatory conditions should reveal whether their causative influence is true and extends beyond coeliac disease.

Table 3 Haplotype analyses of associated SNPs in CYP4F3 (rs1290622 and rs1290625) and CYP4F2 (rs7252046 and rs3093156)

		Frequency		Number			
Gene	*Haplotype*	*Case*	*Control*	*Case*	*Control*	χ^2	*P-value*
CYP4F3	GC	0.888	0.883	539.6	632.2	0.067	0.7960
	GT	0.054	0.082	33	58.4	3.882	0.0488
	AC	0.058	0.035	36	25.4	3.959	0.0466
CYP4F2	TT	0.468	0.417	289	298.3	3.572	0.0587
	AC	0.225	0.288	139	206.5	7.045	0.0079
	AT	0.238	0.239	147.4	171	0.0	0.9861
	TC	0.069	0.056	42.5	40.2	0.922	0.3369

Acknowledgements

We thank all the patients, their physicians, pathologists and the Dutch Coeliac Disease Foundation for participating in this study. We thank Jackie Senior for critically reading the manuscript. This study was supported by grants from the Netherlands Organisation for Scientific Research (Grant 912-02-028), the Dutch Digestive Diseases Foundation (Grants 97-44 and 03-06) and the Coeliac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch Government (Grant BSIK03009). JDR was supported by grants from the American National Institute of Diabetes and Digestive and Kidney Diseases and the Crohn's and Colitis Foundation of America.

References

1 Macdonald TT, Monteleone G: Immunity, inflammation, and allergy in the gut. *Science* 2005; **307**: 1920–1925.
2 Babron MC, Nilsson S, Adamovic S *et al*: Meta and pooled analysis of European coeliac disease data. *Eur J Hum Genet* 2003; **11**: 828–834.
3 Van Belzen MJ, Meijer JW, Sandkuijl LA *et al*: A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* 2003; **125**: 1032–1041.
4 Rioux JD, Silverberg MS, Daly MJ *et al*: Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* 2000; **66**: 1863–1870.
5 van Heel DA, Satsangi J, Carey AH, Jewell DP: Inflammatory bowel disease: progress toward a gene. *Can J Gastroenterol* 2000; **14**: 207–218.
6 Schreiber S, Rosenstiel P, Albrecht M, Hampe J, Krawczak M: Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nat Rev Genet* 2005; **6**: 376–388.
7 Van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD, Lewis CM: Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. *Hum Mol Genet* 2004; **13**: 763–770.
8 Ueda H, Howson JM, Esposito L *et al*: Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 2003; **423**: 506–511.
9 Begovich AB, Carlton VE, Honigberg LA *et al*: A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 2004; **75**: 330–337.
10 Bottini N, Musumeci L, Alonso A *et al*: A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet* 2004; **36**: 337–338.
11 Zhernakova A, Eerligh P, Wijmenga C, Barrera P, Roep BO, Koeleman BP: Differential association of the PTPN22 coding variant with autoimmune diseases in a Dutch population. *Genes Immun* 2005; **6**: 459–461.
12 Cottone M, Marrone C, Casa A *et al*: Familial occurrence of inflammatory bowel disease in celiac disease. *Inflamm Bowel Dis* 2003; **9**: 321–323.
13 Schedel J, Rockmann F, Bongartz T, Woenckhaus M, Scholmerich J, Kullmann F: Association of Crohn's disease and latent celiac disease: a case report and review of the literature. *Int J Colorectal Dis* 2005; **20**: 376–380.
14 Tursi A, Giorgetti GM, Brandimarte G, Elisei W: High prevalence of celiac disease among patients affected by Crohn's disease. *Inflamm Bowel Dis* 2005; **11**: 662–666.
15 Yang A, Chen Y, Scherl E, Neugut AI, Bhagat G, Green PH: Inflammatory bowel disease in patients with celiac disease. *Inflamm Bowel Dis* 2005; **11**: 528–532.
16 Rioux JD, Daly MJ, Silverberg MS *et al*: Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 2001; **29**: 223–228.
17 Monsuur AJ, Bakker PI, Alizadeh BZ *et al*: Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. *Nat Genet* 2005; **37**: 1341–1344.
18 Gabriel SB, Schaffner SF, Nguyen H *et al*: The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225–2229.
19 Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
20 Peltekova VD, Wintle RF, Rubin LA *et al*: Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet* 2004; **36**: 471–475.
21 Tello-Ruiz MK, Curley C, DelMonte T *et al*: Haplotype-based association analysis of 56 functional candidate genes in the IBD6 locus on chromosome 19. *Eur J Hum Genet* 2006; **14**: 780–790.
22 Matter K, Balda MS: Signalling to and from tight junctions. *Nat Rev Mol Cell Biol* 2003; **4**: 225–236.
23 Mankertz J, Tavalali S, Schmitz H *et al*: Expression from the human occludin promoter is affected by tumor necrosis factor alpha and interferon gamma. *J Cell Sci* 2000; **113** (Part 11): 2085–2090.
24 Diosdado B, van Bakel H, Strengman E *et al*: A genomics view on celiac disease points to enhanced neutrophil recruitment and barrier impairment, 2005, (submitted).
25 Helaakoski T, Annunen P, Vuori K, MacNeil IA, Pihlajaniemi T, Kivirikko KI: Cloning, baculovirus expression, and characterization of a second mouse prolyl 4-hydroxylase alpha-subunit isoform: formation of an alpha 2 beta 2 tetramer with the protein disulfide-isomerase/beta subunit. *Proc Natl Acad Sci USA* 1995; **92**: 4427–4431.
26 Christmas P, Jones JP, Patten CJ *et al*: Alternative splicing determines the function of CYP4F3 by switching substrate specificity. *J Biol Chem* 2001; **276**: 38166–38172.

27 Christmas P, Ursino SR, Fox JW, Soberman RJ: Expression of the CYP4F3 gene. Tissue-specific splicing and alternative promoters generate high and low K(m) forms of leukotriene B(4) omega-hydroxylase. *J Biol Chem* 1999; **274**: 21191–21199.
28 Kikuta Y, Kusunose E, Endo K *et al*: A novel form of cytochrome *P*-450 family 4 in human polymorphonuclear leukocytes. cDNA cloning and expression of leukotriene B4 omega-hydroxylase. *J Biol Chem* 1993; **268**: 9376–9380.
29 Kikuta Y, Miyauchi Y, Kusunose E, Kusunose M: Expression and molecular cloning of human liver leukotriene B4 omega-hydroxylase (CYP4F2) gene. *DNA Cell Biol* 1999; **18**: 723–730.
30 Greene TA, Powell P, Nzerem C, Shapiro MJ, Shapiro VS: Cloning and characterization of ALX, an adaptor downstream of CD28. *J Biol Chem* 2003; **278**: 45128–45134.
31 Oda T, Muramatsu MA, Isogai T, Masuho Y, Asano S, Yamashita T: HSH2: a novel SH2 domain-containing adapter protein involved in tyrosine kinase signaling in hematopoietic cells. *Biochem Biophys Res Commun* 2001; **288**: 1078–1086.
32 Altare F, Durandy A, Lammas D *et al*: Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency. *Science* 1998; **280**: 1432–1435.
33 Chua AO, Chizzonite R, Desai BB *et al*: Expression cloning of a human IL-12 receptor component. A new member of the cytokine receptor superfamily with strong homology to gp130. *J Immunol* 1994; **153**: 128–136.
34 Watanabe T, Kitani A, Murray PJ, Strober W: NOD2 is a negative regulator of Toll-like receptor 2-mediated T helper type 1 responses. *Nat Immunol* 2004; **5**: 800–808.
35 Arunachalam B, Phan UT, Geuze HJ, Cresswell P: Enzymatic reduction of disulfide bonds in lysosomes: characterization of a gamma-interferon-inducible lysosomal thiol reductase (GILT). *Proc Natl Acad Sci USA* 2000; **97**: 745–750.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)

Chapter 4.2

Effective detection of HLA risk alleles in celiac disease using tag SNPs

Alienke J Monsuur, Paul I W de Bakker, Dalila Pinto, Willem Verduijn, J Bart A Crusius, Cisca Wijmenga

Submitted

Effective detection of HLA risk alleles in celiac disease using tag SNPs

Alienke J Monsuur[1], Paul I W de Bakker[2], Dalila Pinto[1], Willem Verduijn[3], J Bart A Crusius[4], Cisca Wijmenga[1,5]

[1]Complex Genetics Section, DBG-Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands. [2]Broad Institute of Harvard and M.I.T., Seven Cambridge Center, Cambridge, MA 02141, USA, and Center for Human Genetic Research, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA. [3]Section for Immunogenetics and Transplantation Immunology of the Department of IHB, LUMC, Leiden, The Netherlands. [4]Laboratory of Immunogenetics, Department of Pathology, VU University Medical Centre, Amsterdam, The Netherlands. [5]Genetics Department, University Medical Center Groningen, Groningen, The Netherlands. *Corresponding author:* Prof. Cisca Wijmenga, Genetics Department, room E2.030, University Medical Center Groningen, P.O. Box 30001, 9700 RB Groningen, c.wijmenga@medgen.umcg.nl, c.wijmenga@umcutrecht.nl

Abstract

Background & Aims: The HLA genes, located in the MHC region on chromosome 6p21.3, play an important role in many autoimmune disorders, such as celiac disease (CD), type 1 diabetes (T1D), rheumatoid arthritis, multiple sclerosis, psoriasis and others. Known HLA variants that confer risk to CD, for example, include DQA1*05/DQB1*02 (DQ2.5) and DQA1*03/DQB1*0302 (DQ8). To diagnose the majority of CD patients and to study disease susceptibility and progression, typing these strongly associated HLA risk factors is of utmost importance. However, current genotyping methods for HLA risk factors involve many reactions, and are complicated and expensive. We sought a simple experimental approach using tagging SNPs that predict the CD-associated HLA risk factors.

Methods: We applied a tagging approach that exploits linkage disequilibrium between single nucleotide polymorphism (SNPs) and the CD-associated HLA risk factors DQ2.5 and DQ8 that indicate direct risk, and DQA1*0201/DQB1*0202 (DQ2.2) and DQA1*0505/DQB1*0301 (DQ7) that attribute to the risk of DQ2.5 to CD. In order to evaluate the predictive power of this approach, we performed an empirical comparison of the predicted DQ types, based on six tag SNPs, with those executed with current validated laboratory typing methods of the *HLA-DQA1* and -*DQB1* genes in three large cohorts.

Results & Conclusion: Our results show that this tag SNP method is very accurate and provides an excellent basis for population screening for CD.

Introduction

The HLA genes, located in the major histocompatibility (MHC) region on chromosome 6p21.3, play a role in multiple autoimmune disorders, like celiac disease (CD), type 1 diabetes (T1D), rheumatoid arthritis, multiple sclerosis, psoriasis and others (1-3). The MHC region is highly polymorphic and some genes in this region are involved in multiple disorders. For example, the *HLA-DQA1* and -*DQB1* genes, have alleles that confer risk to both CD and T1D. In most autoimmune diseases not all patients carry the same risk alleles, and multiple risk alleles are likely to be involved (2).

CD, the most common intolerance to a dietary component in Western society, is sustained by an abnormal T cell response to gluten as an environmental factor and is strongly associated with HLA class II genes. Almost 95% of CD patients carry at least one of the two risk molecules DQA1*05/DQB1*02 (i.e. haplotype DQ2.5) and DQA1*03/DQB1*0302 (i.e. haplotype DQ8) (2, 4-7). The molecules encoded by the CD-associated *HLA-DQA1* and -*DQB1* genes form DQα and DQβ heterodimers, which can lead to several functional molecules of which one to four copies can be made. A few variants of these genes predispose to CD (either alone or in combination) when gluten peptides, present in wheat, barley and rye, are presented to CD4+ cells in the lamina propria (8, 9). The most important risk factor for CD is the DQ2.5 haplotype (see Figure 1 and Table 1) (5, 10, 11), with the highest risk in individuals homozygous for this haplotype (8, 12), or those who have a single copy of DQ2.5 and one copy of DQA1*0201/DQB1*0202 (i.e. haplotype DQ2.2) molecules, haplotype DQ8, or DQA1*0505/DQB1*0301 (i.e. haplotype DQ7). The frequency of these alleles in the general population is substantial (>20%), suggesting that these variants are necessary for disease development but not sufficient.

Family-based or population-based screening for the CD risk variants has important diagnostic value in supporting the diagnosis of CD when these alleles are present, and the possibility of CD is minimized when they are not present (they have a high negative predictive value). CD affects almost 1% of the population, although it is estimated that most cases remain undiagnosed (13). Since untreated CD can cause long-term health problems, targeted screening for CD could identify such undiagnosed individuals and prevent life-long symptoms and complications.

Testing for HLA risk molecules is routinely performed using specialized kits, but they often require 24-60 reactions, multiple steps, like amplification and hybridization to a membrane, special software or expertise in analyzing the results and most of these methods are expensive (e.g. DNA PCR-single-strand conformation polymorphism (PCR-SSCP), PCR and sequence specific oligonucleotide probing (PCR-SSOP), PCR-sequence specific primer kits (PCR-SSP, PCR-reverse lineblot (PCR-RLB))

(14-17). Direct typing of the genetic variants that encode the HLA alleles is usually very difficult since most of these variants are surrounded by too many other variants that interfere with primer annealing.

The International HapMap Project and an independent MHC-focused effort (18, 19) have empirically determined the fine-scale patterns of linkage disequilibrium (LD) among local sequence polymorphisms in four population samples. With these resources it is now possible to pick tag SNPs that are in LD with specific HLA variants of interest (i.e. have high r^2 values). Recently, an LD-based tagging approach was shown to predict HLA-DQ2.2 and –DQ2.5 alleles in independent patient samples with a high degree of accuracy (19).

In this study we selected tag SNPs to predict DQ2.2, DQ2.5, DQ7 and DQ8, in three cohorts: CD patients, non-CD trio control families and blood bank controls (HLA typing was available for all individuals). We then examined the sensitivity, specificity, predictive value and the correlation between the SNP-based test and the true HLA variant (r^2). This study represents a first step towards providing a cost-effective population screening method for CD.

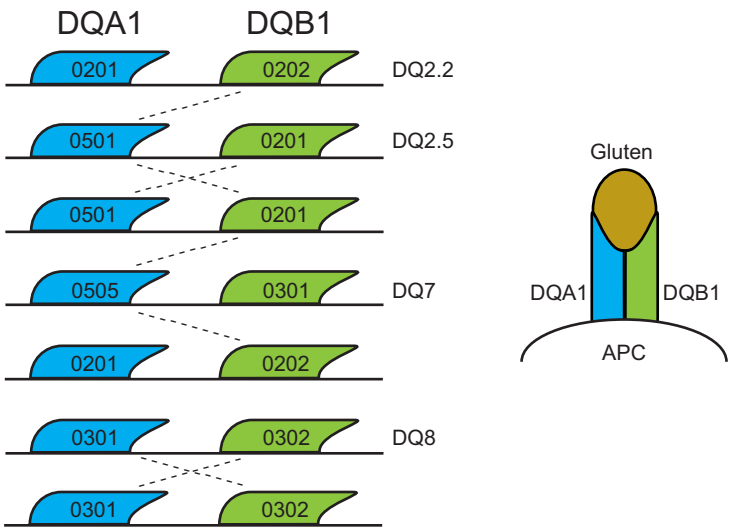


Figure 1. HLA-DQA1* and -DQB1* together form heterodimers of which DQ2.5 and DQ8, either in homozygous or heterozygous state, confer risk to CD due to their ability to present gluten to T cells. DQ2.2 and DQ7 can only confer risk to CD when both are present together or with DQ2.5 (trans effect, see dashed lines). See Table 1b for the possible combinations, the number of risk molecules and the associated risk.

Material and Methods

DNA samples

DNA was available from three different cohorts; these were used to study different aspects of the tag SNP method. The CD cohort had a high number of individuals with HLA-DQ2 risk variants which was useful for testing the positive predictive value. The trio control cohort enabled us to check for Mendelian errors (which were not observed). while the blood bank controls gave a better view of the robustness of the method in the general population. The first cohort consisted of 330 unrelated CD patients of Dutch Caucasian origin (20). Only CD patients diagnosed according to revised ESPGHAN criteria and with a Marsh III lesion confirmed by duodenal biopsy sampling were selected for this study, as described by Van Belzen et al. and Walker-Smith et al. (21-22). A cohort of population-based control trios was derived from families without a history of CD (23). The 86 control trios were selected for the presence of at least one parent carrying haplotype DQ2.5 and were all of Dutch Caucasian origin. HLA typing data was available for 207 of the 264 persons in the 86 trios (see below). The blood bank cohort was part of the ITI two panel (the ITI panel is a DNA panel from the Immunogenetics and Transplantation Immunology Section of the LUMC) and consisted of 219 unrelated, randomly selected, Dutch blood donors. We studied a total of 756 persons. The study was approved by the Medical Ethics Committee of the University Medical Centre Utrecht, and informed consent was obtained from the participants.

HLA typing

The CD cohort and the trio control cohort were typed for *HLA-DQA1* and *-DQB1* genes using a classical PCR-SSCP/heteroduplex method in an official HLA typing laboratory as described elsewhere (12, 14). Full HLA-DQA1 and -DQB1 typing was available for the entire CD cohort. For the trio control cohort full HLA-DQA1 and -DQB1 typing was available for the child and both parents in 35 trios and for the child and one of the parents in 51 trios, leading to a total of 207 persons available for analyses. For the blood bank control cohort, full (four digit) HLA-DRB1, -DQA1 and -DQB1 typing was performed by PCR-SSCP using locally produced and slightly modified primer mixes (24). The typing of this cohort was performed in the European Foundation of Immunogenetics (EFI)-accredited HLA laboratory of the Department of IHB, LUMC, Leiden.

Table 1. DQ molecules and tested tag SNPs

DQ type	DQA1	DQB1	DR	tag SNP	Positive predicting allele(s)	tag SNP	Negative predicting allele
DQ2.2	0201	0202	7	rs2395182, rs7775228	T, G	rs4713586	G
DQ2.5	0501	0201	3	rs2187668	T		
DQ7	0505	0301	5	rs4639334	A		
DQ8	0301	0302	4	rs7454108	G		

a) DQ molecules, the corresponding HLA-DQA1* and -DQB1* alleles, with the DR type and the tag SNPs

DQ molecule 1	DQ molecule 2	Number of functional copies	Genetic risk
DQ2.5	rest	≥1	5.5
DQ2.5	DQ2.5	4	13.1
DQ2.5	no DQ2.2, DQ2.5, DQ7	1	1.3
DQ2.5	no DQ2.5	1-2	2.5
DQ2.5	DQ2.2	2	10.1
DQ2.2 or DQ2.5	rest	1-4	24.4
DQ2.2	DQ7	1	1.8*
DQ2.2	no DQ2.5, DQ7	0	-
DQ7	no DQ2.2, DQ2.5	0	-
DQ2.5	DQ7	2	-
DQ8	rest	1	-
DQ8	DQ8	4	-

b) Combinations of the DQ molecules on the two chromosomes, the number of functional copies and the genetic risk associated to celiac disease (calculated using the CD cohort and the blood bank control cohort). *This risk increases to 4.1 in the DQ2.5 negative group

Tag SNP selection

Tag SNPs were selected that captured the following HLA types: DQ2.2 (2 SNPs for DQ2.2 and one SNP to exclude DQ4 from the DQ2.2 group), DQ2.5 (1 SNP), DQ7 (1 SNP), and DQ8 (1 SNP) (see Table 1). DQ2.5 and DQ8 are risk factors for CD and are carried by ~95% of CD patients (4, 25). The HLA-DQA1*0505 allele of

DQ7 and HLA-DQA1*0501 allele of DQ2.5 only differ by one or a few base-pairs and are thought to have the same functional properties. This also holds for the HLA-DQB1*0202 allele of DQ2.2 and the HLA-DQB1*0201 allele of DQ2.5. Most of the CD patients who do not carry DQ2.5 or DQ8, carry half of the DQ2.5 or DQ2.2 molecule (that is either HLA-DQA1*05 or -DQB1*0202) suggesting that carrying part of the risk molecules has functional implications for the risk of CD (4).

Tag SNP selection was based on genotype data collected in the classical HLA genes and >7,500 common SNP and deletion-insertion polymorphisms across the extended human MHC region (19). We used Tagger (26) to derive SNP-based tests to capture each DQ type in the extended CEU analysis panel (Utah residents with northern and western European ancestry). We first found the SNPs that have the highest r^2 to a DQ type. We proceeded with multi-SNP (haplotype) tests to achieve higher r^2 with which a DQ type was captured (if r^2 <1). For DQ2.2, multiple SNPs were needed that in combination capture this HLA allele. Since there is a lot of variation in the MHC region that can interfere in primer annealing, we only selected SNPs that could be typed using conventional methods (e.g. Taqman).

Tag SNP typing

The tag SNPs were obtained as Assay on Demand (rs2395182, rs4713586, rs4639334) or Assay by Design (rs7454108, rs7775228, rs2187668) from Applied Biosystems (Applied Biosystems, Foster City, California, USA) (see Supplementary Table 1 for assay numbers or primer sequences and their allele labeling). Samples were genotyped using the manufacturer's instructions and analyzed on an ABI PRISM 7900 HT system (Applied Biosystems). All SNPs were typed using the standard amplification protocol as supplied by Applied Biosystems. We obtained end-point measurements for the analysis. Drop-out rates were below 3.57% and are shown in Supplementary Table 1 for each individual SNP. No Mendelian errors were observed for the SNPs in the trio control cohort.

Analyses

The HLA-DQA1 and -DQB1 genotypes as determined at the HLA-typing centers were used to establish the corresponding DQ types (see Figure 1). Due to the high linkage disequilibrium in the MHC region, only a limited set of DQA1*-DQB1* haplotypes (DQ types) are observed in the general population (see http://depts.washington.edu/rhwlab/resMat/dq/linkage.html for an example of common combinations of DQA1* and DQB1* alleles in the Caucasian population), resulting in only a few instances

that did not correspond to canonical DQ types. For the prediction method we inferred DQ types from the tag SNPs. DQ types were determined according to the predicting alleles (see Table 1, e.g. a person was called homozygous DQ8 if rs7454108 was homozygous G, or heterozygous DQ8 if rs7454108 was heterozygous G/A). Only individuals with non-missing data were used for comparing the official typing and the prediction method. DQ types based on the official typing and those from the tag SNP typing method were compared to examine the sensitivity, specificity, positive predictive value (PPV) and r^2.

Results

A total of six SNPs were needed to predict the DQ2.2, DQ2.5, DQ7 and DQ8 risk types for CD. Typing was done in three different cohorts comprising a total of 756 persons (1512 alleles). Drop-out rates for these SNPs were <3.57% as described in Supplementary Table 1. All SNPs were in Hardy-Weinberg equilibrium and no Mendelian errors were observed in the trios.

We observed a high correlation between the three cohorts for the sensitivity, specificity, PPV and the r^2. We grouped the results of the three different cohorts (Table 2) and show the results of each separate group in Supplementary Table 2. Specifications of the individuals of whom the predicted HLA-DQ typing results did not correspond with those from the typing centers are shown in Table 3. For each DQ type we used all persons with non-missing data for the relevant SNPs. A person with missing data for DQ2.5, for example, was excluded from the DQ2.5 analysis, but could be used for the other analyses if genotypes relevant for the other DQ types were present.

At first the sensitivity and specificity for DQ2.2 was high and accurate but the predictive value was low. The SNPs for DQ2.2 (rs2395182, rs7775228) not only tagged DQ2.2 but also included the relatively infrequent DQ4 allele. We therefore decided to tag DQ4 as well (rs4713586) making it possible to call a person DQ2.2 when the alleles were positive for DQ2.2 and negative for DQ4. This led to three tag SNPs being needed for the prediction of DQ2.2, with an overall sensitivity of 0.992, a specificity of 0.995 and a PPV of 0.954. Only seven out of the 1448 tested chromosomes gave false results (0.5%).

The tag SNP selected for prediction of DQ2.5 (rs2187668) showed an overall sensitivity of 0.996, a specificity of 0.994 and a PPV of 0.991. Only seven out of the 1460 tested chromosomes gave false results (0.5%). One out of these seven chromosomes carried half of the DQ2.5 haplotype (DQA1*0501).

The tag SNP for DQ7 (rs4639334) showed an overall sensitivity of 0.968, a

specificity of 0.996 and a PPV of 0.939. Nine out of the 1468 tested chromosomes gave false results (0.6%), of which three carried a rare haplotype consisting of half of the DQ7 haplotype. Two of these three carried the DQB1*0302 of DQ8 on the other half (see results of DQ8 as well).

The tag SNP for DQ8 (rs7454108) showed an overall sensitivity of 0.991, a specificity of 0.995 and a PPV of 0.940. Eight out of the 1486 tested chromosomes gave false results (0.5%). Three of these eight chromosomes carried half of the DQ8 haplotype (DQB1*0302); two of these three chromosomes carried the DQ7 (DQA1*0505) on the other half and were predicted to be both DQ7 and DQ8.

Accepting the prediction of these half haplotypes as good predictions of the risk alleles increases the sensitivity, specificity and PPV slightly.

Discussion

In this study we used a tag SNP approach to predict whether an individual carried the risk DQ types (formed by variants in the *HLA-DQA1* and -*DQB1* genes) that are positively associated with CD. Using this method, only six SNPs were needed to predict the DQ2.2, DQ2.5, DQ7 and DQ8 risk types carried by >95% of CD patients. We determined that for this tagging approach the sensitivity was >0.968, specificity >0.994 and the predictive value >0.940.

Most of the patients without DQ2.5 and DQ8, carried half of the DQ2.5 or DQ2.2 molecule (either HLA-DQA1*05 or -DQB1*0202) suggesting that carrying part of the risk molecules has functional implications for the risk of CD (4, 25). Of our patient group 98.4% carry one of the risk groups (DQ2.2, DQ2.5, DQ7, DQ8 or the DQ types that have half of the risk haplotypes) and 96.1% of all patients were correctly predicted using our method. Overall, the specificity was >0.969 and predictive value was >0.995 when taking into account that some of the false predictions included an allele that is part of a risk haplotype (e.g. the HLA-DQA1*05 allele which is part of the DQ2.5 haplotype).

This method also allowed us to determine whether an individual was homozygous or heterozygous for the risk molecule. Vader et al. demonstrated a >4-fold higher T cell response when gluten was presented by antigen-presenting cells from DQ2 homozygous patients compared to DQ2 heterozygous patients, thereby providing an explanation for the dose-effect of risk molecules for developing CD (9). Al-Toma et al. showed that homozygosity for DQ2.5 was seen more than twice as often in individuals that developed refractory celiac disease and enteropathy-associated T-cell lymphoma, associated with a high morbidity, than in uncomplicated CD (12).

Table 2. Prediction results for combined cohorts

DQ2.2

		+	-			
					sensitivity	0.992
SNP prediction	+	124	6	130	specificity	0.995
	-	1	1317	1318	positive predictive value	0.954
		125	1323	1448	r-squared	0.941

DQ2.5

		+	-			
					sensitivity	0.996
SNP prediction	+	565	5	570	specificity	0.994
	-	2	888	890	positive predictive value	0.991
		567	893	1460	r-squared	0.980

DQ7

		+	-			
					sensitivity	0.968
SNP prediction	+	92	6	98	specificity	0.996
	-	3	1367	1370	positive predictive value	0.939
		95	1373	1468	r-squared	0.903

DQ8

		+	-			
					sensitivity	0.991
SNP prediction	+	109	7	116	specificity	0.995
	-	1	1369	1370	positive predictive value	0.940
		110	1376	1486	r-squared	0.926

Reinton et al. developed a real-time PCR method for detecting CD-associated HLA risk alleles (27). This method requires 11 reactions and even more if homozygous persons for the HLA-risk alleles need to be distinguished from heterozygous persons. It is not clear whether this real-time PCR method can be easily applied to high-throughput typing or not, whereas our method can. We can perform PCR reactions in multiple PCR machines at the same time and use the ABI PRISM 7900 HT system only for end-point measurements. Moreover, Reinton et al. only used a relatively small set of samples to test their method, making it difficult to determine its robustness.

a

Cohort	Prediction of DQ		Chromosome 1				Chromosome 2				Remark
	Allele 1	Allele 2/3	DQA1	DQB1	DQ	DR	DQA1	DQB1	DQ	DR	
1. CD case	DQ2.2	-	0102	0502	DQ5	DR2	03	0301	DQ7*	DR4	DQ2.2 was predicted but not present
2. CD case	DQ2.2	-	0201	02	DQ2.2	DR7	05	0301	DQ7	DR5	DQ2.2 was correctly predicted, DQ7 was present but not predicted
3. Blood bank control	DQ2.2	-	0101	0501	DQ5	DR1	0401	0402	DQ4	DR8	DQ2.2 was predicted but not present
4. Blood bank control	DQ2.2	-	0103	0601	DQ6	DR2	0201	0303	DQ9	DR7	DQ2.2 was predicted but not present
5. CD case	DQ2.2	DQ7	0103	0602/03	DQ6	DR6	05	02	DQ2.5	DR3	DQ2.2 and DQ7 were predicted but not present
6. CD case	DQ2.2	DQ7	0301	0301	DQ7*	DR4	05	02	DQ2.5	DR3	DQ2.2 and DQ7 were predicted but not present
7. Blood bank control	DQ2.2	DQ7	0103	0601	DQ6	DR2	0505	0301	DQ7	DR5	DQ7 was correctly predicted, DQ2.2 was predicted but not present
8. CD case	DQ2.5	-	0102	0502	DQ5	DR2	03	0301	DQ7*	DR4	DQ2.5 was predicted but not present
9. CD case	DQ2.5	-	0201	02	DQ2.2	DR7	05	02	DQ2.5	DR3	DQ2.5 was correctly predicted, DQ2.2 was present but not predicted
10. Trio control	DQ2.5	-	03	0301	DQ7*	DR4	05	0301	DQ7	DR5	DQ2.5 was predicted but not present, DQ7 was present but not predicted
11.CD case	DQ2.5	DQ2.5	0102	0502	DQ5	DR2	03	0301	DQ7*	DR4	DQ2.5 was predicted but not present
12.CD case	DQ2.5	DQ2.5/ DQ8	05	02	DQ2.5	DR3	05	02	DQ2.5	DR3	DQ2.5 was correctly predicted, DQ8 was predicted but not present
13.CD case	DQ2.5	DQ8	05	02	DQ2.5	DR3	0101/02	0602/03	DQ6	DR2	DQ2.5 was correctly predicted, DQ8 was predicted but not present
14.Trio control	DQ2.5	DQ8	05	02	DQ2.5	DR3	01	0501	DQ5	DR1	DQ2.5 was correctly predicted, DQ8 was predicted but not present
15. Blood bank control	DQ7	DQ7	05	0301	DQ7	DR5	0104	0503	DQ5	DR6	DQ7 was twice predicted, but only present once
16. Blood bank control	DQ7	DQ8	0505	0301	DQ7	DR5	0103	0603	DQ6	DR6	DQ7 was correctly predicted, DQ8 was predicted but not present
17. CD case	-	-	0102	0602/03	DQ6	DR2	0301	0302	DQ8	DR4	DQ8 was present but not predicted
18. Trio control	DQ8	-	03	0302	DQ8	DR4	05	0301	DQ7	DR5	DQ8 was correctly predicted, DQ7 was present but not predicted

Table 3. False-positive and false-negative results. DQ types, predicted using the tag SNPs are shown as well as the official HLA-DQA1 and -DQB1 typing of both chromosomes, with the corresponding DQ and DR types. a) Wrongly predicted DQ types; b) wrongly predicted results from individuals who carry part of the risk haplotypes for CD (in rare DQA1*/DQB1*combinations) and who therefore contributed to the CD risk. DQX are rare combinations for which there is no generally used DQ name. *Only DQ7/DR5 gives risk to CD, not DQ7/DR4

b

Cohort	Prediction of DQ			Chromosome 1*				Chromosome 2*				Remark
	Allele 1	Allele 2	Allele 3	DQA1	DQB1	DQ	DR	DQA1	DQB1	DQ	DR	
19. CD case	DQ2.5	DQ2.5		05	02	DQ2.5	DR3	05	0602/03	DQX		One of the DQ2.5 alleles was correctly predicted, the other was predicted but not present, although half of the haplotype was present in a rare combination
20. CD case	DQ2.5	DQ7		05	02	DQ2.5	DR3	05	0503	DQX		DQ2.5 was correctly predicted, DQ7/DR5 was predicted but not present, although half of the haplotype was present in a rare combination
21. CD case	DQ2.5	DQ7	DQ8	05	02	DQ2.5	DR3	05	0302	DQX		DQ2.5 was correctly predicted, DQ7/DR5 and 8 were predicted but not present due to the presence of a rare combination of a half DQ7 and a half DQ8 haplotype
22. Trio control	DQ2.5	DQ7	DQ8	05	02	DQ2.5	DR3	05	0302	DQX		DQ2.5 was correctly predicted, DQ7/DR5 and 8 were predicted but not present due to the presence of a rare combination of a half DQ7 and a half DQ8 haplotype
23. CD case	DQ8	-		0401/0601	0302	DQX		0401	0402	DQ4	DR8	DQ8 was predicted but not present, although half of the DQ8 haplotype was present in a rare combination

De Bakker et al. showed two examples that used the tagging method for CD and systemic lupus erythematosus (19). They chose two SNPs to capture DQ2.2 and DQ2.5 in the same CD cohort (N=330) that we have used in this paper.
The rs4988889(T), rs2858331(C) haplotype was used to determine the presence of DQ2.2 and the rs4988889(T), rs2858331(T) haplotype was used to determine DQ2.5. Although the SNPs look promising in determining DQ2.5 homozygosity or DQ2.2/DQ2.5 heterozygosity, it was often difficult to distinguish DQ2.2/X heterozygous from the DQ2.5/X heterozygous individuals (X is any other allele excluding DQ2.2 or DQ2.5), due to phase uncertainty of the alleles at the two SNPs. An individual who is heterozygous for rs4988889 (G/T) has one copy of DQ2.2 or DQ2.5. When he/she is also heterozygous for rs2858331 (C/T), then it is uncertain which of these alleles (either C or T) is on the same chromosome as the T allele of rs498889, and therefore forms either DQ2.2 or DQ2.5. In contrast to these examples are the SNPs we used in the current study, which are very capable of determining whether an individual is homozygous for DQ2.2 or DQ2.5, heterozygous for DQ2.2 or DQ2.5 or not possessing the DQ2.2 or DQ2.5 haplotype at all.

This method is attractive because it is cost-effective and the experimental procedures are straightforward, using routine genotyping equipment. Although more work is needed to validate this approach for diagnostic purposes, this work provides a foundation for simple SNP-based population screening for CD. Although population screening has been discussed for a long time and it would certainly be helpful in finding all the undiagnosed CD patients, which could prevent negative outcomes (13), no steps have actually been taken towards implementing it largely due to the cost of classical HLA typing and the need to repeat serology tests during an individual's life time. Since an individual can develop antibodies to CD later in life, repeated testing would impose an extra burden on people who might not be at any risk. It would be easy, cheap and quick to use our tag SNP method to determine which part of the population (~25%) needs to be screened more extensively for CD. As this test requires very little DNA and is somewhat insensitive to DNA quality, it can also be used with DNA material isolated from e.g. biopsies, whole-genome amplified DNA, and DNA isolated from FTA cards. Furthermore, this test determines which individuals are not at risk for developing CD and who therefore do not need further serology tests.

We have described a robust method to predict the risk DQ types involved in CD with high accuracy. This method can also be applied to T1D, in which DQ2.5 and DQ8 are also known risk factors, or more generally for other immune-related diseases with known HLA risk alleles.

Acknowledgements

We thank all healthy individuals, the patients, their physicians, and the Dutch Celiac Disease Foundation for participating in this study. We thank Harry van Someren as database manager and Jackie Senior for critically reading the manuscript. This study was supported by grants from the Netherlands Organization for Scientific Research (grants 902-22-094 and 912-02-028), the Dutch Digestive Disease Foundation (grants 97-44 and 03-06) and the Celiac Disease Consortium, an innovative cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch government (grant BSIK03009).

Conflicts of Interests: No conflicts of interest exist
There is no potential investigator conflict of interest

References

1. Ghodke Y, Joshi K, Chopra A, Patwardhan B. HLA and disease. Eur J Epidemiol 2005;20:475-88.
2. Jones EY, Fugger L, Strominger JL, Siebold C. MHC class II proteins and disease: a structural perspective. Nat Rev Immunol 2006;6:271-82.
3. Lie BA, Thorsby E. Several genes in the extended human MHC contribute to predisposition to autoimmune diseases. Curr Opin Immunol 2005;17:526-31.
4. Karell K, Louka AS, Moodie SJ, Ascher H, Clot F, Greco L, Ciclitira PJ, Sollid LM, Partanen J. 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.
5. Lundin KE, Scott H, Fausa O, Thorsby E, Sollid LM. T cells from the small intestinal mucosa of a DR4, DQ7/DR4, DQ8 celiac disease patient preferentially recognize gliadin when presented by DQ8. Hum Immunol 1994;41:285-91.
6. Lundin KE, Scott H, Hansen T, Paulsen G, Halstensen TS, Fausa O, Thorsby E, Sollid LM. Gliadin-specific, HLA-DQ(alpha 1*0501,beta 1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. J Exp Med 1993;178:187-96.
7. Vader LW, de Ru A, van der Wal Y, Kooy YMC, Benckhuijsen W, Mearin ML, Drijfhout JW, van Veelen P, Koning F. Specificity of Tissue Transglutaminase Explains Cereal Toxicity in Celiac Disease. J. Exp. Med. 2002;195:643-649.
8. Koning F. Celiac disease: caught between a rock and a hard place. Gastroenterology 2005;129:1294-301.
9. Vader W, Stepniak D, Kooy Y, Mearin L, Thompson A, van Rood JJ, Spaenij L, Koning F. The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. PNAS 2003;100:12390-12395.
10. Margaritte-Jeannin P, Babron MC, Bourgey M, Louka AS, Clot F, Percopo S, Coto I, Hugot JP, Ascher H, Sollid LM, Greco L, Clerget-Darpoux F. HLA-DQ relative risks for coeliac

disease in European populations: a study of the European Genetics Cluster on Coeliac Disease. Tissue Antigens 2004;63:562-7.
11. Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E. Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. J Exp Med 1989;169:345-50.
12. Al-Toma A, Goerres MS, Meijer JW, Pena AS, Crusius JB, Mulder CJ. Human leukocyte antigen-DQ2 homozygosity and the development of refractory celiac disease and enteropathy-associated T-cell lymphoma. Clin Gastroenterol Hepatol 2006;4:315-9.
13. Mearin ML, Ivarsson A, Dickey W. Coeliac disease: is it time for mass screening? Best Pract Res Clin Gastroenterol 2005;19:441-52.
14. Carrington M, Miller T, White M, Gerrard B, Stewart C, Dean M, Mann D. Typing of HLA-DQA1 and DQB1 using DNA single-strand conformation polymorphism. Hum Immunol 1992;33:208-12.
15. Ronningen KS, Spurkland A, Iwe T, Vartdal F, Thorsby E. Distribution of HLA-DRB1, -DQA1 and -DQB1 alleles and DQA1-DQB1 genotypes among Norwegian patients with insulin-dependent diabetes mellitus. Tissue Antigens 1991;37:105-11.
16. Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. Tissue Antigens 1992;39:225-35.
17. Buyse I, Decorte R, Baens M, Cuppens H, Semana G, Emonds MP, Marynen P, Cassiman JJ. Rapid DNA typing of class II HLA antigens using the polymerase chain reaction and reverse dot blot hybridization. Tissue Antigens 1993;41:1-14.
18. A haplotype map of the human genome. Nature 2005;437:1299-320.
19. De Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke X, Monsuur AJ, Whittaker P, Delgado M, Morrison J, Richardson A, Walsh EC, Gao X, Galver L, Hart J, Hafler DA, Pericak-Vance M, Todd JA, Daly MJ, Trowsdale J, Wijmenga C, Vyse TJ, Beck S, Murray SS, Carrington M, Gregory S, Deloukas P, Rioux JD. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet 2006;38:1166-72.
20. Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, Franke L, van 't Slot R, van Belzen MJ, Lavrijsen IC, Diosdado B, Daly MJ, Mulder CJ, Mearin ML, Meijer JW, Meijer GA, van Oort E, Wapenaar MC, Koeleman BP, Wijmenga C. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. Nat Genet 2005;37:1341-4.
21. Van Belzen MJ, Meijer JW, Sandkuijl LA, Bardoel AF, Mulder CJ, Pearson PL, Houwen RH, Wijmenga C. A major non-HLA locus in celiac disease maps to chromosome 19. Gastroenterology 2003;125:1032-41.
22. Walker-Smith J, Guandalini S, Schmitz J, Shmerling D, JK V. 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.
23. Van Belzen MJ, Koeleman BP, Crusius JB, Meijer JW, Bardoel AF, Pearson PL, Sandkuijl LA, Houwen RH, Wijmenga C. Defining the contribution of the HLA region to cis DQ2-positive coeliac disease patients. Genes Immun 2004;5:215-20.
24. Verduyn W, Anholts JD, Versluis LF, Parlevliet J, Drabbels J, De Meester J, Tilanus

MG, Doxiadis, II, Giphart MJ, Schreuder GM. Six newly identified HLA-DRB alleles: DRB1*1121, *1419, *1420, *1421, DRB3*0203 and DRB5*0103. Tissue Antigens 1996;48:80-6.

25. Louka AS, Sollid LM. HLA in coeliac disease: unravelling the complex genetics of a complex disorder. Tissue Antigens 2003;61:105-17.
26. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet 2005;37:1217-23.
27. Reinton N, Helgheim A, Shegarfi H, Moghaddam A. A one-step real-time PCR assay for detection of DQA1*05, DQB1*02 and DQB1*0302 to aid diagnosis of celiac disease. J Immunol Methods 2006;316:125-32.

Part 5

General discussion

This discussion is built up of three parts and a section on the future perspectives. The first part is about complex genetics in general, the second looks at the gene *myosin IXB* (*MYO9B*), and the third looks at what we have learned from the work described here in relation to celiac disease. The discussion closes by looking at the future perspectives. Several terms that might not be familiar to most readers are explained in boxes.

- **Complex genetics in general**
  - Complex genetic disorders and challenges in finding genes
  - The pros and cons of studying linkage disequilibrium
  - Pinpointing the causal variant
  - Stratification
  - Overlap in genetic variants between different disorders
- ***MYO9B***
  - Genetic studies
  - Location of the association
  - Sequencing of *MYO9B*
  - Which hypothesis for *MYO9B* is the most plausible for celiac disease?
- **Celiac disease in general**
  - What have we learned from the genetic studies on celiac disease?
  - Future perspectives
  - Back to the aim of this thesis

Complex genetics in general

Complex genetic disorders and challenges in finding genes

Complex genetic disorders occur due to a combination of variations in several genes as well as in environmental factors (see Figure 1). The changes in the genes are probably subtle and in order to develop a disorder a person may have to have a combination of several such variants. Environmental factors are needed to trigger and/or sustain the disorder, like gluten in celiac disease. It is expected that each complex genetic disorder involves 10-50 genes and that a person needs to have variants in 10 of these genes for example in order to develop the disorder. Based on this we do not expect all patients to have the same set of genetic variants (genetic heterogeneity) and healthy persons may well carry some of these genetic variants, but not enough or not the right

combination to develop the disorder. A difference in ethnic background can lead to different associated genes, while some genetic differences may be due to an ancestor who introduced a specific disease variant into a population. Some of the genes involved may carry a high risk for developing the disorder, but most genes will only carry a low relative risk. Different disorders that are related biologically, or seen in one person or family, may have a shared genetic background (see Figure 1 and some of the examples discussed below). There may be a set of genes that leads to a general vulnerability for intestinal disorders for example, while other genes may specifically orchestrate this vulnerability towards a specific disorder, like celiac disease or ulcerative colitis.

The genetic heterogeneity and the involvement of multiple low-risk genes make it difficult to find the genes involved in these disorders. On the other hand, finding a gene involved in celiac disease might well lead to the observation that this gene is also involved in other gastrointestinal or inflammatory disorders.

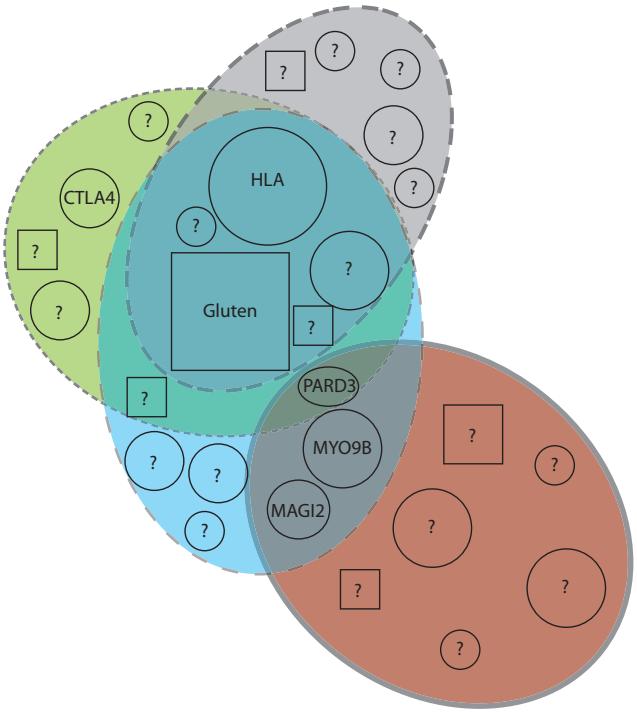


Figure 1. Complex genetic disorders. Celiac disease and other complex genetic disorders occur due to a combination of variations in several genes (circles) as well as in environmental factors (squares).Heterogeneity between patients of one disorder and overlap between disorders is seen. Large transparent circles with dashed lines represent celiac disease populations; the large transparent circle with the solid line represents the ulcerative colitis population.

The pros and cons of studying linkage disequilibrium

There are millions of single nucleotide polymorphisms (SNPs) and other variations, like repeats in the genome that we can use to look for association of a gene to a disorder.

We do not have to test them all in order to find these genes and variants, since we can tag SNPs (see Box 1). Together with the new techniques for testing a huge number of SNPs at the same time (e.g. 300,000 SNPs) this has led to the possibility of screening the whole genome for genes involved in a certain disorder.

The drawback of this method is that it excludes a huge number of SNPs based on estimates of their linkage disequilibrium (LD) to others, while in reality, or in our specific Dutch population, the LD might be slightly different and important SNPs will therefore be missed (4). The SNPs are mostly chosen such as to cover 80% of the genetic information on the genome, but due to the failure of some SNPs in practice and to small differences in LD, the real coverage of the genome is less than 80%, so that false-negative results are obtained.

Although LD provides the possibility of screening the whole genome relatively easily, it also has some drawbacks. It is possible that a disease-associated SNP is found in a region of high LD (the so-called block). It is then difficult to pinpoint which of the SNPs in the block is the actual cause of the disorder, for example, in IBD a 250 kb region was found with five genes on chromosome 5q31 but we still do not know which gene is the one truly associated with IBD (5, 6).

Box 1

Linkage disequilibrium (LD)

Many SNPs are not independent and cosegregate to the next generation. The combination of two neighbouring SNPs, each with two alleles will then not result in four possible haplotypes but only two or three, depending on the degree of dependence or disequilibrium. LD is expressed in several measures, like D' and r^2, with the latter used most widely. The higher the r^2 the more the SNPs are in disequilibrium (values range from 0-1). When two SNPs have an $r^2 = 1$, only one (the so-called tag SNP) is needed for genetic testing, since it predicts perfectly the outcome of the other SNP (1).

HapMap

HapMap is a database that contains most of the SNPs known so far, tested in up to four different populations. www.hapmap.org (2)

Tagger

Program developed by Paul de Bakker, which uses the HapMap information to exclude redundant SNPs and select the SNPs that need to be tested, based on r^2 measures. www.broad.mit.edu/mpg/tagger (3)

Pinpointing the causal variant

Whereas in monogenetic disorders just one variant or mutation in a gene leads to the disorder, complex genetic disorders are influenced by subtle changes in many genes. If we find a disease-related gene, we do not automatically expect to find a non-synonymous SNP (amino acid changing SNP) that leads to a truncated protein, which is what is seen for almost 60% of the monogenetic disorders (7). We know that besides non-synonymous SNPs, synonymous SNPs (no amino acid changed), intronic SNPs and SNPs outside a gene can also play a role. SNPs can influence splice sites, or influence regulatory elements, the latter are often unidentified but if they are identified, it is not always obvious which gene(s) is regulated by this splice site.

An example of a synonymous SNP influencing a disorder was found in the MDR1 gene, which did not change mRNA or protein levels, but did change the folding of the protein, due to the presence of a rare codon that might change the time needed to produce the protein (8). Much bioinformatics and functional work, taking many years, is needed to identify which variant is involved and how it influences a gene.

Stratification

Genetic studies are now mostly done using trio families or case-control cohorts (see Box 2). For trio families the transmission disequilibrium test is mostly used. The advantage of this method is that the control alleles come from exactly the same genetic background as the case alleles, thereby excluding the possibility of stratification. The drawbacks of this method are that three persons need to be tested instead of two (normally only a case and a control) and the power of this method is lower and cannot easily be increased. The lower power is due to the difficulty of obtaining DNA from both the parents of the affected person, so that the number of trio families is often not very high. A second reason for the low power is that a SNP only has two alleles, and if a parent is homozygous for a certain allele, then it is not obvious which allele was transmitted to the child and which allele was not transmitted, leading to a loss of countable transmissions. Although the control allele matches perfectly with the case allele, one should keep in mind that different ethnic populations should not be mixed. A recent IBD study of the chromosome 5q31 region showed that the observed association did not come from the Ashkenazi Jews in the study cohort, which diluted the results found in the North American group (6).

A more widely used method is the case-control design. Single cases can be more easily obtained and the number of controls can be easily increased by, for example, taking a large cohort of blood bank donors, or by sharing control cohorts among

Box 2

Trio studies

A trio family consists of an affected child with both parents. Using the transmission disequilibrium test (TDT) method, the alleles of the parents transmitted to the affected child are seen as case alleles and the alleles that are not transmitted are seen as control alleles.

Case-control studies

A group of individual cases (one from each family) is compared to a group of individual controls that are either selected for not having the disorder or that are randomly collected when the disorder has a low population frequency (e.g. less than 1%). For each SNP the allele frequency of the case group is compared to the allele frequency of the control group, often using a chi-square test in a 2x2 table.

Linkage region

Studies in families with two or more affected children can locate regions on the chromosomes that are likely to harbour a gene involved in that disorder.

research groups. This is now being done by the Wellcome Trust Case Control Consortium (WTCCC) for several genome-wide screens being carried out in England; they are all using the same control cohort of 3000 persons in order to reduce the costs of such projects. The screens will search for genetic signposts for tuberculosis, coronary heart disease, type 1 diabetes, type 2 diabetes, rheumatoid arthritis, Crohn's disease and ulcerative colitis, bipolar disorder and hypertension (http://www.wtccc.org.uk). Although case-control studies give a high power to find association, there is a chance that stratification will be present. If the control group comes from a different genetic background than the case group, the observed association might not be due to association of a gene to disease but to stratification. However, if such studies avoid using isolated populations, it is unlikely that stratification will be present in a country like the Netherlands. In genome-wide screens there are also enough markers tested to screen for possible stratification.

DLG5 is a gene associated with IBD, but this finding could not be replicated in all the populations tested (9-12). Tenesa et al. showed that the cases were more similar in allele frequency than the controls, suggesting either differences in population frequency of the controls or perhaps stratified control groups (10). The first explanation means that one should be very careful when grouping several studied populations into a meta–analysis, while the second is a warning to look for the possibility of stratification.

There are various methods to check for stratification between cases and controls, which can, when necessary, exclude the controls so that the groups are no longer stratified. Our finding of association between celiac disease and *MYO9B* in a Dutch population was debated in the literature and stratification of our control group was suggested. We therefore checked for stratification using two methods: Genomic Control Method and the program STRUCTURE (De Kovel et al., unpublished observations) (13-15). The genomic control method looks at the inflation of χ^2-values in cases and controls relative to a neutral expectation, and STRUCTURE is used to search for subpopulations. The 112 SNPs tested resided in candidate genes and that could explain why the χ^2-values were somewhat higher then expected (1.039 instead of 1), but it hardly affected the results when we corrected the observed p-values for it. Using STRUCTURE we were not able to identify any subpopulations. A discussion on the reasons why *MYO9B* has not been replicated in other populations can be found in section *Myosin IXB*, Genetic studies.

Overlap in genetic variants between different disorders

As stated earlier, we expect to see an overlap in the genetic backgrounds of related disorders (Figure 1). Two well-known examples of a shared genetic background in immune-related disorders are *PTPN22* and *CTLA4* (16-20). *PTPN22* is involved in type I diabetes, rheumatoid arthritis and Graves thyroiditis, but not in celiac disease. *CTLA4* is involved in type I diabetes, autoimmune thyroid disease and celiac disease.

The common biological features of related disorders and their overlapping linkage regions can guide our thinking as to what genes would be good candidate genes for these disorders. In the work described in this thesis we initially searched for genes involved in celiac disease, but later extended this search to inflammatory bowel disorders (IBD), which are made up of Crohn's disease and ulcerative colitis. We found that the gene *MYO9B* on chromosome 19p13.1 is involved in celiac disease (chapter 3.1 and discussed below in the section on *MYO9B*) (21). Based on the overlap of some pathological features between celiac disease and ulcerative colitis, the co-occurrence in families and individuals, and the overlap in their linkage regions (see box 2) we decided to look for association of *MYO9B* with IBD (5, 22-29). Cooperating with two other research groups has enabled us to show that *MYO9B* is also associated to ulcerative colitis (chapter 3.4) (30). In chapter 4.1 we describe how we selected immunological candidate genes in the overlapping linkage regions for celiac disease and IBD on chromosome 19p13.1 (linkage regions called MYO9B and IBD6) and on chromosome 5q31 (CELIAC2 and IBD5), but we found no association of these genes

with celiac disease (31). John Rioux's group performed a search in their IBD cohort for 56 candidate genes in the IBD6 region, including the genes we tested (chapter 4.1), but they too found no association (32). In chapter 3.5 we describe experiments to look at genes involved in tight junctions in relation to celiac disease, Crohn's disease and ulcerative colitis; here we found that *PARD3* and *MAGI2* were associated to both celiac disease and ulcerative colitis (Wapenaar et al. submitted). Two of our studies on common genes between inflammatory disorders confirmed our hypothesis for a shared genetic background.

One difficulty with this genetic overlap is that it is not known whether the control groups are also affected by the disorder being studied. We assume that this is not a major problem if a disorder is uncommon, e.g. with a maximum of 1% of the general population (and therefore the controls) being affected. But if a particular gene is involved in multiple disorders, this might well be a problem, since then 1-5% of the controls could be affected and this would lead to false-negative results. However, for most of our studies we used a control cohort of blood bank donors, who in general are healthier than the average population since patients of chronic disorders do not donate blood.

Myosin IXB

Genetic studies

In 2003 our group found linkage for celiac disease to a region on chromosome 19 (27). The microsatellite marker that showed the highest linkage was also associated to celiac disease in a follow-up study, and was located in intron 1 of the *MYO9B* gene. SNP analysis in this gene confirmed this association (chapter 3.1) (21). Since the screening with microsatellite markers could have missed association somewhere else in this linkage region, we decided to re-screen the whole region using a dense set of SNPs that were selected using Tagger. However, except for the known association to *MYO9B*, we could not find any other association in this region; we thus excluded the possibility that the linkage in this region was due to another gene. In our study we used two independent case-control cohorts and therefore replicated our own findings. Three of the six SNPs tested in our associated haplotype block were needed to tag the four haplotypes that account for over 97% of all the observed haplotypes (these SNPs were rs2305767, rs1457092 and rs2305764). Logistic regression showed that SNP rs2305764 alone could completely explain the observed association, but the two

other SNPs cannot be statistically excluded and all three are considered to be markers for risk of celiac disease. It is not yet clear which of these three SNPs, or other SNPs in the haplotype block, is causal for celiac disease.

Several groups tried to replicate our finding, but all except one failed (33-37) (Sanchez et al., in press) (see Table 1). There are several possible reasons why replication of our study is proving so difficult: the other studies could be false-negatives, there could be differences in LD structure, the mutation in *MYO9B* is only found in the Dutch population, or our study could be a false-positive (type I error).

False-negatives

The observed odds ratio (OR) gives a value to the strength of the finding but it may be inflated. In reality, the true OR could also approach the lower limit of the confidence interval, as has often been seen in meta-analyses of several genetic disorders (38). If so, then the power of most of the replication studies was probably too low – see Table 2 for a power calculation given the observed ORs and their 95% confidence intervals. Taking into account that the ORs in the replication populations were closer to the lower confidence interval limit, very large numbers of cases and controls would be needed for a valid replication.

Differences in LD structure

Differences in LD structure may also influence the chance of replication. The associated SNPs in our study have, to date, not been proven to be causative, they seem rather to be a marker of disease. The LD between these SNPs and the real mutation might differ between populations, thereby diminishing the power of the replication studies. The studies by Hunt et al., Amundsen et al. and Sanchez et al. typed the three tag SNPs, and the studies by Cirillo et al. and Giordano et al. typed parts of one or two tag SNPs, respectively (Sanchez et al., in press) (33-35, 37). If the LD is slightly different in the populations there might have been better tag SNPs for these studies that could have led to a positive association. In our IBD study on *MYO9B* we typed rs2305767 and although we found association to ulcerative colitis in our Dutch cohort we could not replicate this in the other cohorts. If we had only used this SNP, then we would have missed the associations on the two other tag SNPs and on the coding SNP, and it would have been doubtful whether *MYO9B* was really associated to ulcerative colitis. By using a comprehensive set of SNPs we were able to replicate the association and identify the coding SNP as the one most associated with ulcerative colitis.

Table 1. Overview of the genetic studies on MYO9B (21, 30, 33-35, 37, 39) (Sanchez et al., in press, Wolters et al., in preparation). a) celiac disease, b) refractory celiac disease, c) Inflammatory Bowel Disorders, d) systemic lupus erythematosis, e) rheumatoid arthritis, *case and control allele frequencies were based on the transmitted and untransmitted alleles of the trios, **rs962917 in the Canadian/Italian population

	Reference	Country of origin	Number of		Frequency (case/control (%))													
					rs7259292			rs2305767			rs1545620			rs1457092**			rs2305764		
			Ca	Co	Ca	Co	p-value	Ca	Co	p-value	Ca	Co	p-value	Ca	Co	p-value	Ca	Co	p-value
a	Monsuur et al.	Dutch	463	686	3	3.2	NS	38.4	46.1	0.0003	41.5	33.8	0.0002	38.6	30.4	3.2x10-5	45.7	35.7	2.1x10-6
	Hunt et al.	UK	375	1366				41.8	40.9	NS				35.4	37.7	NS	38.7	41.5	NS
	Amundsen et al.	Norwegian/Swedish	326	562				39	40.1	NS				37.6	34.1	NS	41.6	42.6	NS
	Giordano et al.	Italian	252 trios*											39	40	NS	41	40	NS
	Sanchez et al.	Spanish	90	345				37	46	0.03				37	33	NS	43	37	NS
	Cirillo et al.	South Italian	223	600													40.1	43.5	NS
b	Wolters et al.	Dutch	62	1624	10	3.2	0.00002	35	44	0.05	45	36	0.04	37	33	NS	40	38	NS
c	Van Bodegraven et al.	Dutch	IBD 588	1624	2.7	2.8	NS	38.3	44.1	0.0006	40.8	35.8	0.0028	38.2	33	0.0013	42.6	38.1	0.0075
			CD 298	1624	3.1	2.8	NS	40.2	44.1	NS	40.9	35.8	0.019	38.3	33	0.013	41.2	38.1	NS
			UC 290	1624	2.2	2.8	NS	36.3	44.1	0.0006	40.6	35.8	0.028	38.2	33	0.016	44.1	38.1	0.0073
		UK	IBD 1315	2371	3.3	3.4	NS	40.1	41.1	NS	44.1	40.2	0.0016	40.8	36.9	0.0013	44	40.8	0.01
			CD 735	2371	3.5	3.4	NS	41.5	41.1	0.063	43.3	40.2	0.038	39.5	36.9	0.075	42.7	40.8	NS
			UC 580	2371	3	3.4	NS	38.3	41.1	NS	45	40.2	0.0035	42.5	36.9	0.0008	45.5	40.8	0.0039
		Canadian/Italian	IBD 814	445	6	5.9	NS	39.6	42.1	NS	46.2	41.7	0.034	40.4	36.9	0.09	41.7	39.1	NS
			CD 164	445	5.5	5.9	NS	42.3	42.1	NS	42.7	41.7	NS	38	36.9	NS	38.4	39.1	NS
			UC 650	445	6.1	5.9	NS	38.9	42.1	NS	47.1	41.7	0.015	41	36.9	0.056	42.5	39.1	NS
		Combined analysis	IBD 2717	4440						0.002			1.9x10-6			2.1x10-6			0.00014
	Amundsen et al.	Norwegian/Swedish	IBD 457	562				39.7	40.3	NS				36.7	35.6	NS	42	43.1	NS
			CD 149	562				42	40.3	NS				36.2	35.6	NS	41.4	43.1	NS
			UC 308	562				38.6	40.3	NS				36.9	35.6	NS	42.2	43.1	NS
d	Sanchez et al.	Spanish	356	345				44	46	NS				41	33	0.002	44	37	0.01
e	Sanchez et al.	Spanish	349	345				42	46	NS				39	33	0.01	41	37	NS

Table 2. Number of cases needed to achieve 80% power given the odds ratios found in Monsuur et al. and van Bodegraven et al. (21, 30).

	Odds Ratio	Lower limit of 95% CI*	Higher limit of 95% CI*	Reference
Allelic Odds Ratio	1.5	1.12	1.87	Monsuur et al.
Number of cases	472	5400	218	
Allelic Odds Ratio	1.23	1.12	1.35	Van Bodegraven et al.
Number of cases	1647	5400	826	
Genotypic Odds Ratio (AG/AA)	1.66/2.27	1.23/1.56	2.13/3.3	Monsuur et al.
Number of cases	139	481	69	

Confidence Interval (CI). Calculations were performed using the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html). High risk allele frequency and marker allele frequency were set to 0.4, prevalence to 0.01 and control/case ratio to 2.

Heterogeneity and founder mutations

Heterogeneity in celiac disease could also be the cause of the lack of replication, which could be reflected by the fact that the allele frequencies in the control populations differ. If this gene plays a large role in the Dutch population and only a small role in other populations, then the chance of replication is indeed very small. Consequently if the frequency of the disease mutation in the control populations of these other studies is higher than in the Dutch population the differences between cases and controls will be lower. It is general knowledge that allele frequencies of many SNPs differ between ethnic populations due to global migrations. A mutation could have been brought to a part of Europe and the population with this mutation could have expanded into the Dutch population resulting in our observed association. Or a founder of the Dutch population could have acquired a mutation in *MYO9B*. In our group, a study was performed that looked at refractory celiac disease (RCD) patients, type 2 (Wolters et al., in preparation). Besides some weak association to two of the SNPs used in the original *MYO9B* paper, we found association to a low frequency SNP, rs7259292. This SNP did not differ between cases and controls (3% in each population) but in the RCD group the frequency increased to 10%. This could mean there are several mutations to be found.

False-positive?

After several studies were unable to replicate our *MYO9B* association to celiac disease, two studies assembled all the case and control alleles in one table, showing that the cases from several populations showed more or less equal allele frequencies, whereas the Dutch control group differed in allele frequency from the other control groups, suggesting a type I error (due to stratification) (35, 37). We do not agree that this is the reason for the non-replication; there could be differences in allele frequencies between populations. The association of *DLG5* to IBD was a similar situation where the cases were more alike than the controls (9-12). Later on, Sanchez et al. found association for *MYO9B* to celiac disease, systemic lupus erythematosus and rheumatoid arthritis and their control frequencies were similar to our Dutch controls (Sanchez et al., in press). The Dutch study also showed the association in two different cohorts and the gene is located directly under the top of the Dutch linkage peak on chromosome 19. It has a plausible function in relation to CD. We did, however, agree that given the number of negative replications we needed to investigate the possibility of stratification in our cohort. This was done in a study by De Kovel et al. (discussed above) that looked at inflation of χ^2-values in cases and controls relative to the neutral expectation and searched for subpopulations; neither of these factors were found. However, it has been shown that stratification is difficult to find, so we cannot exclude this completely.

Several points should be added to the story of *MYO9B* to provide a more balanced picture. We found association of *MYO9B* to ulcerative colitis (30). We used the same control group as in the celiac disease study, which could explain the observed association to ulcerative colitis if it is stratified. However, we added a set of extra blood bank controls to increase the number (from 686 up to 1624 controls), thereby decreasing the chance that the control group was stratified. In addition, we collaborated with two other groups that tested the associated SNPs in their own laboratories on their own case and control groups. All the groups found association for *MYO9B* to ulcerative colitis, so there is now one initial finding and two replication studies. One other published study by Amundsen et al. could not find any association of *MYO9B* to IBD (see Table 2) (39); they also reported calculations for ulcerative colitis and Crohn's disease separately but still observed no significant differences. They calculated if they had 80% power to observe association based on the OR found for celiac disease using the associated haplotype and the AA genotype. These calculations were done using the whole IBD cohort, whereas we only found association to ulcerative colitis but not to Crohn's disease, so their actual power is lower. Van Bodegraven et al. showed that the OR in ulcerative colitis was lower than the one reported for celiac disease

(it is generally seen that first studies find a higher OR than subsequent replication studies (38)) (see Table 1 for the ORs reported by Van Bodegraven et al.). And lastly, the power is mostly calculated such that there is an 80% chance of observing true association, leaving a 20% chance that the association was present but not detected. The lack of obvious stratification, the finding and replication of association of *MYO9B* to ulcerative colitis, and our hypothesis that there is a shared genetic background between related disorders strengthens our findings for celiac disease.

Location of the association

In the study in which we used celiac disease patients, we found association in a block that started close to exon 14 and ended just after *MYO9B* (see also Figure 1 of chapter 1.2, association starts before the IQ domains and ends after the gene). We assumed in our Nature Genetics paper that differences in the RhoGAP domain lead to the association to celiac disease, since our most associated SNP is close to the RhoGAP domain. However, regression analysis did not show which SNP is the causative variant so all the domains, starting from the IQ domains, could be altered.

If we look at the ulcerative colitis group, the association was most strong in rs1545620, which is a nonsynonymous variant in exon 20 leading to an amino acid change (Ala1011Ser) in the third calmodulin binding IQ domain of *MYO9B*. The study by Wolters et al. on the refractory type II celiac patients showed association to rs7259292, which is located before exon 14, deep in the intron. The newest HapMap data contain more SNPs and gives a more precise definition of the block structure (see also Figure 1 of chapter 3.4) (30). We can now see that although all the SNPs are in one region/block with an $r^2 > 0.5$, using a more stringent definition shows that the most associated SNPs in these studies are all on other blocks (rs7259292, rs154562 and rs2305764).

It is quite likely that there are different mutations involved in celiac disease, ulcerative colitis and refractory type II celiac disease patients. If this also holds true for the celiac disease patients, it might explain the difficulty in replicating the association we found for *MYO9B*.

Sequencing of *MYO9B*

We wanted to elucidate which variant is causative for the association of *MYO9B* to celiac disease so we sequenced the associated region in *MYO9B* (Trynka et al. in preparation).

We have found that the haplotype that consisted of the A alleles of our associated SNPs (at-risk haplotype) was more frequent in cases than in controls, and vice versa, the haplotype that consisted of the G alleles (non-risk haplotype) was most frequent in controls (21). Using cases and controls matched for the at-risk haplotype means that the variants found in the case group are most probably causative and not neutral (40). This selection minimizes the chance of the detected variants being benign or neutral. In other words, this study design aims to rule out all variants that are less interesting; the group of variants that is different between the matched case-controls could potentially be causative variants.

We used 10 patients and 8 controls (4 HLA-DQ2 positive and 4 HLA-DQ2 negative persons) homozygous for the at-risk haplotype block, as well as 2 controls (HLA-DQ2 negative) who carried the most frequent non-risk haplotype. The controls carrying the non-risk haplotype should match the closest to the reference sequence in the databases; they were used for quality control. Four of the controls that carried the risk-haplotype were also matched for being DQ2-positive, since they carry the high-risk HLA-DQ2 type but have not become celiac disease patients this suggests that they might have fewer risk variants than other controls, so the chance that they may carry a MYO9B risk variant is lower. All variants that are different compared to the non-risk haplotype are interesting, and their influence on the protein should be looked at using bioinformatic tools.

The 3' part of *MYO9B* was sequenced, in total 30 kb, encompassing exon 14 to the 3'UTR. This included the whole 19 kb LD block that we reported to be associated

Box 3

Indel
: Insertion or deletion of one or more base pairs relative to the reference DNA strand

Transition
: A type of point mutation in which one purine or pyrimidine is replaced by another base of the same type, e.g. A->G and C->T

Transversion
: A type of base pair substitution in which a purine base is replaced by a pyrimidine base or vice versa, e.g. A->T and G->C

Phylogenetic analysis of SNPs
: To decipher the historical relationships among SNPs

in the Nature Genetics paper (see chapter 3.1) (21). So far we have found 93 variants of which 78 are SNPs and 15 indels (Box 1) (Figure 2). Of the 78 SNPs 24 were transversions and 54 were transitions. At least six of the SNPs had a higher frequency in cases with the at-risk haplotype compared to the controls with the at-risk haplotype. We are currently studying these six SNPs in our case-control cohort and we are looking at the effect of the variants, e.g. whether they influence the splicing of MYO9B, and whether they are non-synonymous SNPs or synonymous SNPs. Phylogenetic analysis could also be used to help to find the causal variant; this technique was reported in a paper on asthma (41). Using this technique we would hope to find, within the risk-haplotype, the sub haplotypes that are most distinctive between cases and controls and to discover which variants are on the sub haplotype in the cases that dsitnguishes them from the controls. Such variants could well be causative

At the same time we are trying to find different transcripts of MYO9B in peripheral blood and biopsies of the small intestine, we have already observed some small changes; i.e. exon 14 and 36 are not always present in the transcripts. These exons are not located in domains and the consequence of these splice variants is not clear. The differences found so far cannot be correlated to disease status, since both case and control biopsies have shown these differences.

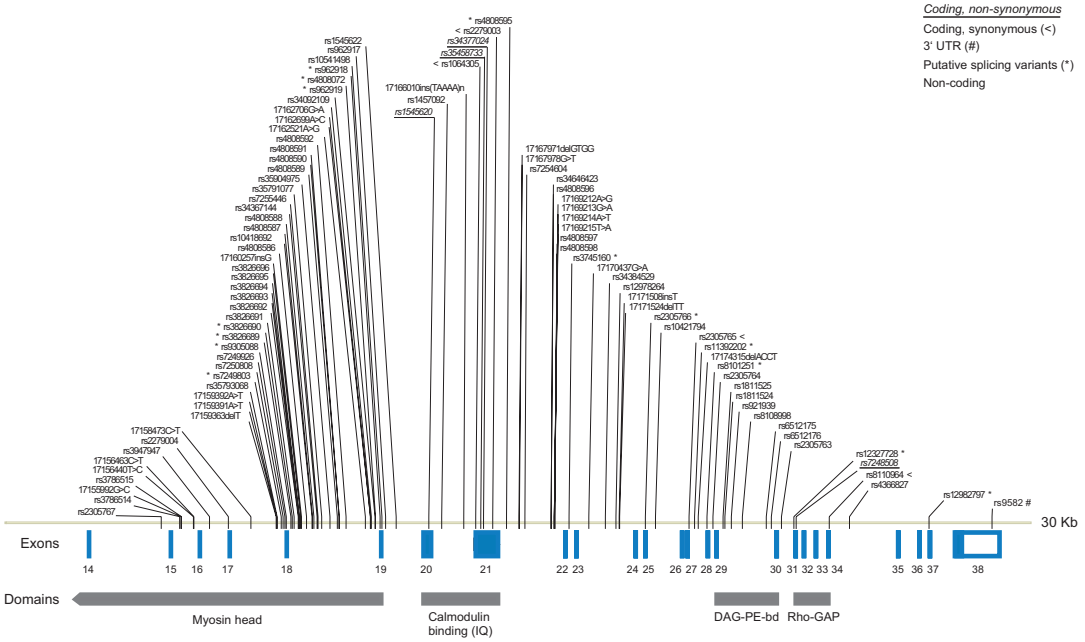


Figure 2. Location of sequence variants in *MYO9B*

Which hypothesis for MYO9B is the most plausible for celiac disease?

Several hypotheses were mentioned in the introduction on MYO9B (chapter 1.2) but the results described in this thesis mean one hypothesis is more plausible than the others.

Celiac disease and ulcerative colitis are disorders that show overlap in patients, families and linkage regions, suggesting a shared genetic background (5, 22-29). In both disorders the mucosa of the intestine are damaged and there is an increased permeability of the intestinal barrier (42). This increase in permeability is not only confined to active celiac disease patients (patients not yet on a gluten-free diet) and untreated ulcerative colitis patients, but also to patients who are being treated and some first-degree relatives of celiac disease patients. This suggests that the increased permeability is due to alterations in some genes. In celiac disease patients we also see that there are many proliferating cells in the epithelial barrier that do not lead to a normal crypt-villous structure of the small intestine. This could be due to more open tight junctions and resulting less contact inhibition that would lead to the switch from proliferating to differentiating cells.

We have found association of *MYO9B*, *PARD3* and *MAGI2* not only to celiac disease but also to ulcerative colitis (21, 30). PARD3 and MAGI2 are both genes involved in the tight junctions of the intestinal epithelial barrier (see chapter 3.5). MYO9B could also be involved in tight junctions via its influence on rho, which is involved in actin dynamics and tight junction functioning (43).

Based on this evidence and reasoning, I consider that the first hypothesis of chapter 1.2 is the most plausible: "*MYO9B is able to affect rho kinases and therefore the control of tight junction functioning and cytoskeletal modifications, leading to an increased permeability of the intestinal barrier.*"

Future studies will need to elucidate the exact role of MYO9B and there are several options:

- performing expression studies (with and without knockdowns of MYO9B) to determine in which cells MYO9B plays an important role, and what effect it has on these cells,
- locating which splice fragments are present in different cell types and if there are differences between healthy and affected individuals,
- measuring the trans epithelial resistance of the enterocytes in normal and MYO9B knockdown cell layers,
- determining which route gluten takes into the body in healthy and affected individuals,

- performing mice studies with and without a knock-out of MYO9B to see if repair of small wounds in the enterocyte layer is altered,
- determining how strong the influence of MYO9B on rho is and whether there is a different competition of proteins like BIG1 with rho for binding to MYO9B on comparing healthy individuals to affected individuals.

This list is not complete and proper step-wise planning is needed to determine which steps need to be taken first. It would help if the exact genetic mutation that causes the association of *MYO9B* to celiac disease is known.

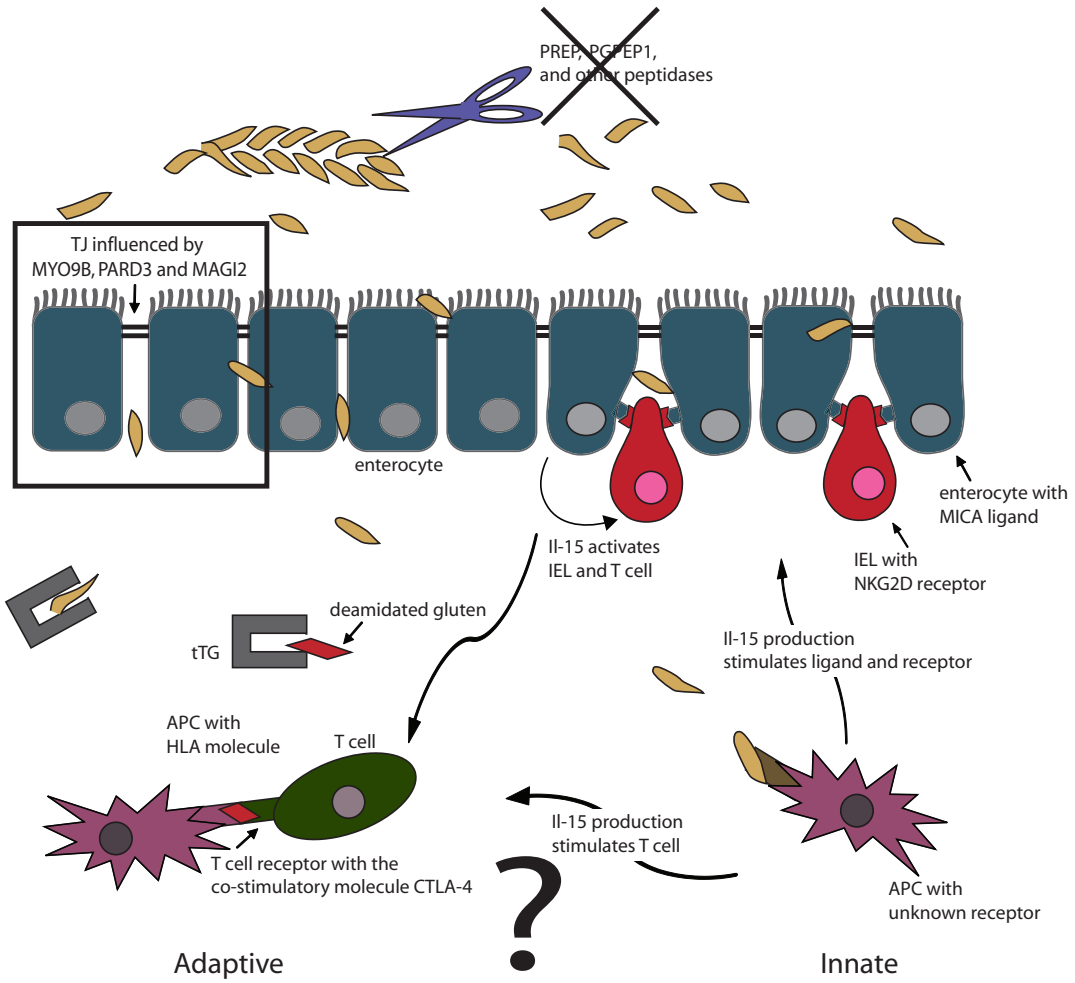


Figure 3. Overview of the outcomes of the genetic studies in this thesis in relation to celiac disease.

Celiac disease in general

What have we learned from the genetic studies on celiac disease? (see Figure 3)

No genetic involvement is seen for genes able to breakdown gluten (see chapter 2.1 and 2.2, studies on *PREP* and *PGPEPI*) (44, 45). We can, of course, not exclude that some genes involved in gluten breakdown have been overlooked due to their unknown functions. Although *PREP* is not genetically involved, its biological involvement in celiac disease must not be underestimated, and it is a good candidate for therapeutic options (46).

We have found association of *MYO9*B to celiac disease and ulcerative colitis (21, 30). We performed the study for ulcerative colitis and *MYO9B* assuming that there is altered permeability of the intestinal barrier in celiac disease and IBD patients and that MYO9B is involved in this alteration. This hypothesis is also supported by the biological evidence. However, our hypothesis on the role of MYO9B is not yet proven. Since we still do not know what MYO9B's role is, we cannot overlook the possibility that the biological aspects underlying the shared genetics lie more in the field of altered inflammatory control by lymphocytes.

We, and several other groups, studied the genes involved in inflammatory pathways, however, besides the involvement of *CTLA-4*, there was no evidence that inflammatory genes are involved in celiac disease. On the one hand, we can say that we should stop looking for association in this group of genes since we have not found anything so far. On the other hand, we did not test all of the genes and small effects may also have been overlooked. I would not recommend investigating these genes by candidate gene studies, but the new possibility of genome-wide screening will include all genes and could show us whether inflammatory genes do in fact play a role.

Recently, in combination with our group, David van Heel's group performed a genome-wide study (van Heel et al., submitted) using a British cohort of celiac disease patients and some of the controls from the WTCCC study. Besides the HLA region, they found a second locus that showed strong association. This region is a long block containing four genes, two of which are involved in inflammation. The most associated SNP was replicated in our Dutch cohort and in a Scottish cohort. We need to study this region in order to find the causative variant as well as performing functional studies. Our first guess is that the causative gene is one of the inflammatory genes. This would suggest that the inflammatory pathways are more involved in celiac disease then we thought based on my own research.

MYO9B, an example of a hypothesis-free approach that led to a functionally interesting candidate and an interesting pathway

There are over 20,000 genes in the human genome, some of which have a known function. Several studies have looked at genes based on their biological function in relation to disease, for example, the studies presented in chapter 2.1 and 2.2. The difficulty with these functional candidate studies is that there are so many genes that could fit with the observed clinical features of a disorder. We can then easily overlook genes or pathways that harbor the disease-causing mutations. *MYO9B* was not at first an obvious candidate (and in terms of size and functional studies it still isn't) but is a good example of how hypothesis-free methods, like linkage studies and case-control studies that screen a region or the whole genome, can lead to a functional candidate, which in turn guides the thinking of gene finding in other disorders (see chapter 4.4 and 4.5), or of interesting pathways that can be explored (see chapter 4.5).

The *HLA* genes that were found over 30 years ago contribute 40% of the disease risk in the population, *MYO9B* accounts for some 25% and there is a minor effect from the *PARD3*, *MAGI2* and *CTLA4* genes, suggesting that more than 75% of the population attributable risk can be explained. However, if *MYO9B* only plays a role in the Dutch celiac disease population, there is more room for other genetic factors in other populations.

Future perspectives

There is still much to discover and learn when it comes to complex genetic disorders. In the period that this research was being performed, there were reports on knowledge about the block structure in the human genome, that junk DNA is not junk since it can contain all kind of signals, RNAi, the role of synonymous SNPs, and much more. Since the discovery of DNA over 50 years ago, we have discovered different layers of information in the DNA and it would not be surprising if there are more layers to find. With respect to the known functions of the DNA and the genetics of celiac disease, there are several aspects that need to be improved in order to get better results.

- We need a better understanding of the functional role of the susceptibility genes in celiac disease.
- We need to gain a better understanding of the strength of the role of each susceptibility gene in celiac disease and of the genetic interaction of these genes, by developing tools that can look at these interactions, without running into the problem of multiple testing and statistical flukes. We also need to know if a certain

gene is only involved in a subset of patients.

- A larger patient group is needed as well as a larger control cohort that can be stratified for more factors than just age and gender, thus better standardized information is needed from the physicians (which is already standard in the IBD field).
- It would be good if there is a standard set of replication populations (case-controls and trio families) each with a large enough power, not only within Europe but also in each country that performs such genetic studies. There are enough patients in each country (~1:100), so this should be feasible.
- We need to develop better tools to find the causative variant underlying association, perhaps by developing a standard strategy after the association study:
  - Replicate the findings in other cohorts
  - Perform stratification on phenotypic subtypes of the disorders and on other associated genes
  - Sequence the different haplotypes in the associated region (all haplotypes not just risk/non-risk) in large groups of patients
  - Phylogenetic analysis of the sequencing results
  - Automated database searches for the influence of the found variants
  - Expression studies, splicing studies and other functional studies.

Back to the aim of this thesis

"The aim of this thesis was therefore to find some of the genes underlying the susceptibility to celiac disease. We hoped these would lead to a better understanding of the mechanisms involved in celiac disease and reveal more about the complex genetic background of celiac disease and other related disorders." (see Preface and Aim of the thesis)

We excluded biological candidate genes, like *PREP* and *PGPEP1*, and found association to *MYO9B*, *PARD3* and *MAGI2*, but we expect there are many more genes still to find. *MYO9B* put us on the track of the permeability of the intestinal barrier and we learned much about this aspect of celiac disease. We revealed more about the complex genetic background and found that it is indeed "complex", especially since finding association to *MYO9B* has raised questions about the actual mutation, its role in different populations, and its functional role. We found genes shared by different disorders and learned that the complexity of these disorders does not stop at the boundaries proposed by physicians on the basis of patients' visible symptoms, but that there are underlying biological aspects that are shared between disorders.

References

1. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nat Genet. 2001 Oct;29(2):229-32.
2. A haplotype map of the human genome. Nature. 2005 Oct 27;437(7063):1299-320.
3. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet. 2005 Nov;37(11):1217-23.
4. de Bakker PI, Burtt NP, Graham RR, Guiducci C, Yelensky R, Drake JA, et al. Transferability of tag SNPs in genetic association studies in multiple populations. Nat Genet. 2006 Nov;38(11):1298-303.
5. Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, et al. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet. 2001 Oct;29(2):223-8.
6. Silverberg MS, Duerr RH, Brant SR, Bromfield G, Datta LW, Jani N, et al. Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease. Eur J Hum Genet. 2007 Mar;15(3):328-35.
7. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet. 2003 Mar;33 Suppl:228-37.
8. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science. 2007 Jan 26;315(5811):525-8.
9. Stoll M, Corneliussen B, Costello CM, Waetzig GH, Mellgard B, Koch WA, et al. Genetic variation in DLG5 is associated with inflammatory bowel disease. Nat Genet. 2004 May;36(5):476-80.
10. Tenesa A, Noble C, Satsangi J, Dunlop M. Association of DLG5 and inflammatory bowel disease across populations. Eur J Hum Genet. 2006 Mar;14(3):259-60; author reply 60-1.
11. Daly MJ, Pearce AV, Farwell L, Fisher SA, Latiano A, Prescott NJ, et al. Association of DLG5 R30Q variant with inflammatory bowel disease. Eur J Hum Genet. 2005 Jul;13(7):835-9.
12. Noble CL, Nimmo ER, Drummond H, Smith L, Arnott ID, Satsangi J. DLG5 variants do not influence susceptibility to inflammatory bowel disease in the Scottish population. Gut. 2005 Oct;54(10):1416-20.
13. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999 Dec;55(4):997-1004.
14. Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet. 2005 Nov;37(11):1243-6.
15. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000 Jun;155(2):945-59.
16. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet. 2004 Apr;36(4):337-8.
17. Zhernakova A, Eerligh P, Wijmenga C, Barrera P, Roep BO, Koeleman BP. Differential

association of the PTPN22 coding variant with autoimmune diseases in a Dutch population. Genes Immun. 2005 Sep;6(6):459-61.
18. van Belzen MJ, Mulder CJ, Zhernakova A, Pearson PL, Houwen RH, Wijmenga C. CTLA4 +49 A/G and CT60 polymorphisms in Dutch coeliac disease patients. Eur J Hum Genet. 2004 Sep;12(9):782-5.
19. van Heel DA, Hunt K, Greco L, Wijmenga C. Genetics in coeliac disease. Best Pract Res Clin Gastroenterol. 2005 Jun;19(3):323-39.
20. Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature. 2003 May 29;423(6939):506-11.
21. Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, et al. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. Nat Genet. 2005 Dec;37(12):1341-4.
22. Cottone M, Marrone C, Casa A, Oliva L, Orlando A, Calabrese E, et al. Familial occurrence of inflammatory bowel disease in celiac disease. Inflamm Bowel Dis. 2003 Sep;9(5):321-3.
23. Schedel J, Rockmann F, Bongartz T, Woenckhaus M, Scholmerich J, Kullmann F. Association of Crohn's disease and latent celiac disease: a case report and review of the literature. Int J Colorectal Dis. 2005 Jul;20(4):376-80.
24. Tursi A, Giorgetti GM, Brandimarte G, Elisei W. High prevalence of celiac disease among patients affected by Crohn's disease. Inflamm Bowel Dis. 2005 Jul;11(7):662-6.
25. Yang A, Chen Y, Scherl E, Neugut AI, Bhagat G, Green PH. Inflammatory bowel disease in patients with celiac disease. Inflamm Bowel Dis. 2005 Jun;11(6):528-32.
26. Babron MC, Nilsson S, Adamovic S, Naluai AT, Wahlstrom J, Ascher H, et al. Meta and pooled analysis of European coeliac disease data. Eur J Hum Genet. 2003 Nov;11(11):828-34.
27. Van Belzen MJ, Meijer JW, Sandkuijl LA, Bardoel AF, Mulder CJ, Pearson PL, et al. A major non-HLA locus in celiac disease maps to chromosome 19. Gastroenterology. 2003 Oct;125(4):1032-41.
28. Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, McLeod RS, Griffiths AM, et al. Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. Am J Hum Genet. 2000 Jun;66(6):1863-70.
29. van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD, Lewis CM. Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. Hum Mol Genet. 2004 Apr 1;13(7):763-70.
30. van Bodegraven AA, Curley CR, Hunt KA, Monsuur AJ, Linskens RK, Onnie CM, et al. Genetic variation in myosin IXB is associated with ulcerative colitis. Gastroenterology. 2006 Dec;131(6):1768-74.
31. Curley CR, Monsuur AJ, Wapenaar MC, Rioux JD, Wijmenga C. A functional candidate screen for coeliac disease genes. Eur J Hum Genet. 2006 Nov;14(11):1215-22.
32. Tello-Ruiz MK, Curley C, DelMonte T, Giallourakis C, Kirby A, Miller K, et al. Haplotype-based association analysis of 56 functional candidate genes in the IBD6 locus on chromosome 19. Eur J Hum Genet. 2006 Jun;14(6):780-90.
33. Hunt KA, Monsuur AJ, McArdle W, Kumar PJ, Travis SP, Walters JR, et al. Lack of association of MYO9B genetic variants with coeliac disease in a British cohort. Gut. 2006 July;55(7):969-72.

34. Amundsen SS, Monsuur AJ, Wapenaar MC, Lie BA, Ek J, Gudjonsdottir AH, et al. Association Analysis of MYO9B Gene Polymorphisms with Celiac Disease in a Swedish/ Norwegian Cohort. Human Immunology. 2006;67(4-5):341-5.
35. Cirillo G, Di Domenico MR, Corsi I, Gagliardo T, Del Giudice EM, Perrone L, et al. Do MYO9B genetic variants predispose to coeliac disease? An association study in a cohort of South Italian children. Dig Liver Dis. 2007 Jan 29.
36. Nunez C, Marquez A, Varade J, Martinez A, Polanco I, Maluenda C, et al. No evidence of association of the MYO9B polymorphisms with celiac disease in the Spanish population. Tissue Antigens. 2006 Dec;68(6):489-92.
37. Giordano M, Marano C, Mellai M, Limongelli MG, Bolognesi E, Clerget-Darpoux F, et al. A family-based study does not confirm the association of MYO9B with celiac disease in the Italian population. Genes Immun. 2006 Oct;7(7):606-8.
38. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet. 2001 Nov;29(3):306-9.
39. Amundsen SS, Vatn M, Wijmenga C, Sollid LM, Lie BA. Association analysis of MYO9B gene polymorphisms and inflammatory bowel disease in a Norwegian cohort. Tissue Antigens. 2006 Sep;68(3):249-52.
40. Spencer DH, Bubb KL, Olson MV. Detecting disease-causing mutations in the human genome by haplotype matching. Am J Hum Genet. 2006 Nov;79(5):958-64.
41. Laitinen T, Polvi A, Rydman P, Vendelin J, Pulkkinen V, Salmikangas P, et al. Characterization of a common susceptibility locus for asthma-related traits. Science. 2004 Apr 9;304(5668):300-4.
42. Macdonald TT, Monteleone G. Immunity, inflammation, and allergy in the gut. Science. 2005 Mar 25;307(5717):1920-5.
43. Soldati T. Unconventional myosins, actin dynamics and endocytosis: a menage a trois? Traffic. 2003 Jun;4(6):358-66.
44. Diosdado B, Stepniak DT, Monsuur AJ, Franke L, Wapenaar MC, Mearin ML, et al. No genetic association of the human prolyl endopeptidase gene in the Dutch celiac disease population. Am J Physiol Gastrointest Liver Physiol. 2005 Sep;289(3):G495-500.
45. Monsuur AJ, Stepniak D, Diosdado B, Wapenaar MC, Mearin ML, Koning F, et al. Genetic and functional analysis of pyroglutamyl-peptidase I in coeliac disease. Eur J Gastroenterol Hepatol. 2006 Jun;18(6):637-44.
46. Stepniak D, Spaenij-Dekking L, Mitea C, Moester M, de Ru A, Baak-Pablo R, et al. Highly efficient gluten degradation with a newly identified prolyl endoprotease: implications for celiac disease. Am J Physiol Gastrointest Liver Physiol. 2006 May 11.

Summary

Celiac disease (CD) is a complex, multifactorial genetic disorder that affects around 1% of the population. A genetic susceptibility to celiac disease means that celiac patients cannot consume a normal Western diet since gluten proteins, which are present in wheat, barley and rye, are the main environmental factor provoking this disorder. In patients who eat gluten the immune system responds as if it is a threat to the body and the response destroys the normal surface of the small intestine. This surface has a huge surface area due to its microscopic folding and it is normally capable of absorbing all the necessary nutrients from ingested food. However, in celiac patients who eat gluten the intestinal surface is severely reduced by the loss of the microscopic folds and they are unable to absorb sufficient nutrients. The patient thus suffers from a variety of symptoms related to malnutrition. The symptoms of CD were long thought to arise only from the intestinal damage, but we now know that other organs are also affected. The only treatment is a life-long gluten-free diet, which allows the intestinal surface to largely recover.

Besides this environmental factor, variations in genes are also involved in the disorder. These include common variations that are also present in the general population, but CD patients have either a bad combination or too many of these variants and become ill. The *HLA-DQA1* and *-DQB1* genes together form a protein that is used by the immune system to recognize and present invaders like bacteria and viruses to other immune cells that are capable of eliminating them. These genes contain several variations, leading to a protein that can differ in the kind of bacteria and viruses that it presents. Some of the variants form the DQ2.5 and DQ8 proteins, which have a special capability; they are able to present gluten to the cells of the immune system. If gluten is presented in this way, the body's immune system will respond as if the gluten is an invader. Almost all CD patients have either the DQ2.5 or DQ8 variant and will become ill when they eat gluten. However, over 25% of the general population also has these variants and these people do not develop CD. Therefore the DQ2.5 or DQ8 variant is not sufficient to develop the disorder. CD patients probably carry other gene variants that together with DQ2.5 or DQ8 and lead to CD. The search for some of these genes and variants was the main research topic of this thesis, but it is likely that there are many more still to be identified.

Chapter 1.1 gives an overview of the clinical and genetic aspects of CD. The description of our search for genes is divided into three parts in this thesis (parts 2, 3 and 4) since each part covers genes involved at different locations in the intestine.

In Part 2 we look into two enzymes that are involved in breaking down gluten in the lumen of the intestine.

Gluten proteins are difficult to digest as most of the enzymes that cut food proteins into small parts for digestion are not capable of cutting gluten. We thought that CD patients might have variants of the specific gluten-digesting enzymes that are not efficient in digesting gluten. We searched for differences between patients and healthy controls in two of these enzymes, pyroglutamyl-peptidase I (PGPEP1) and prolyl endopeptidase (PREP) (see chapters 2.1 and 2.2) but could not find any differences in the genes, nor in the function of the proteins that were formed. Although we could not find concrete evidence that these genes are involved in the susceptibility for CD, we should not underestimate their role in CD. One of the enzymes, PREP, can also be found in certain bacteria. It is currently being investigated whether adding this enzyme to a patient's diet can help cut the gluten into such short parts before it enters the small intestine that the immune system no longer responds to the proteins.

In Part 3 we describe our search for genes involved in the increased permeability of the intestinal barrier in CD.

After looking into the lumen of the small intestine, we also searched for genes present in the cell barrier that separates the body from the lumen of the intestine. Our search was not initially aimed at this barrier but was directed to this by the results that we present in chapter 3.1. We searched in the linkage region on 19p13.1 (named MYO9B, previously called CELIAC4) for a gene involved in CD. Using an initial case-control cohort we found association to the gene *myosin IXB* (*MYO9B*) for which we already had some evidence and which was located directly under the linkage peak. We then used a second cohort to replicate these findings and obtained a p-value of $2.1*10^{-6}$ for the combined cohorts. The allelic odds ratio (OR) for *MYO9B* was 1.56 and individuals homozygous for the at-risk allele had a 2.3-times higher risk of developing CD. *MYO9B* can explain around 25% of the population attributable risk, which can be added to the 40% explained by the *HLA-DQA1* and -*DQB1* genes. Not a lot is known about the function of MYO9B but, based on the literature and on our knowledge of CD, we postulate that the gene is involved in increasing the permeability of the intestinal barrier (see also chapter 1.2 and the discussion in Part 5 for more information on the *MYO9B* gene).

Finding this gene expanded our understanding of CD in several ways. *MYO9B* was the first gene to be found since the discovery of the *HLA* genes involved in CD (besides some weak evidence for the *CTLA4* gene). Initially, most research focused on looking

for genes with a function that could be interesting for the disorder, but with the chance of overlooking genes with unknown functions or genes that had functions that we had never considered interesting for CD. *MYO9B* was found using a genome-wide approach that did not consider the function. However, the identification of *MYO9B* broadened our thinking and led to our hypothesis that MYO9B is involved in increasing the intestinal barrier's permeability. We are still looking for confirmation that *MYO9B* is also involved in CD in other populations, as can be seen in chapters 3.2 and 3.3. With our new knowledge on *MYO9B* and on complex genetic disorders in general, we decided to see whether *MYO9B* was also involved in other intestinal disorders and this work led to identifying *MYO9B* as a factor in ulcerative colitis (p-value = $1.9*10^{-6}$; OR = 1.2). This was confirmed in other populations and is described in chapter 3.4. How MYO9B is involved in increasing the permeability of the intestinal barrier is unknown, but it may have an influence on the tight junctions. This made us wonder if genes that form the tight junctions or direct the function of the tight junction in the intestinal barrier were involved in CD and inflammatory bowel disorders (IBD; a joint name for Crohn's disease and ulcerative colitis). We were able to show that two of these genes, *MAGI2* and *PARD3*, play a role in susceptibility for CD and ulcerative colitis; this work is described in chapter 3.5 (results in CD for *PARD3*: p-value = $3.2*10^{-5}$, OR 1.23, and for *MAGI2*: p-value = 0.0018; OR 1.26). Thus we have now found genes involved in physical functions that we had not previously considered, and we were able to show a shared genetic background between CD and ulcerative colitis.

In Part 4 we describe searching for genes involved in the inflammatory component of CD and looking for a good method to screen for the HLA risk factors.

Some intestinal disorders show an overlap between linkage regions and an overlap between the biological pathways involved. We decided to use this information in our search for associated genes. The immune system plays a role in both CD and IBD and these disorders share some linkage regions. We searched the whole 250 kb region on chromosome 5q31, which has been shown to be associated to IBD, and also chose several immune-related genes in the region on chromosome 19, but we could not find a convincing association (see chapter 4.1). However, in the meantime, we had found the *MYO9B* gene in this region.

Lastly we decided to improve the diagnostic testing methods for the known *HLA* variants. These genes are located in a region with a huge number of variants,

which makes testing for the exact variants difficult and screening for them (to aid diagnosing CD or excluding the possibility of CD) difficult and expensive. We knew that there were other variants in the region that are simpler to test, and that some of these variants are always inherited together, with DQ2.5 for example, the so-called linkage disequilibrium. We searched for these variants and tested if one or more of these variants could predict precisely if a person has DQ2.5. We also did this for DQ2.2, DQ7 and DQ8 which are all variants of the *HLA-DQA1* and *-DQB1* genes that alone, or in combination, can yield susceptibility to CD. Using this method, we found that only six SNPs were needed to predict the DQ2.2, DQ2.5, DQ7 and DQ8 risk types carried by >95% of CD patients. For this tagging approach we determined that the sensitivity was >0.968, specificity >0.994 and the predictive value >0.940. Our method is simple and cost effective, so it can be used for screening relatives of CD patients or even whole populations.

Part 5 is a discussion of what we have learnt from the work done for this thesis.

The discussion of this thesis focuses on three parts: the search for genes involved in complex genetic disorders, the role of MYO9B in CD, and what we have learnt from our genetics studies in CD. It ends with some future perspectives on research into CD and other complex genetic disorders in general.

The search for genes involved in CD is difficult since we do not expect to find mutations, which disrupt the proteins, rather genetic variations that alter the protein functions and which are present in the normal population as well. Besides this celiac patients can differ from one another in the kinds of variants they have (genetic heterogeneity). The use of linkage disequilibrium can help in detecting the variants without having to test all the variants in the genome, but linkage disequilibrium can also mask which of the gene variants is the true susceptibility variant. As a result we still do not know which variant in *MYO9B* is the true susceptibility variant. The variants that are being used for genetic testing can vary in allelic frequency between populations, resulting in false-positive and false-negative results when the case and control groups are not carefully matched. We have discussed the difficulty in replicating the association of *MYO9B* to CD, and reasoned that a combination of genetic heterogeneity, not testing the actual susceptibility variant, and cohorts that were too small could be the causes of the negative replication studies.

Our most important finding was that the *MYO9B* gene is involved in the genetics of CD and ulcerative colitis. We also showed that two tight junction genes, *MAGI2* and *PARD3*, are involved in CD and ulcerative colitis. We hypothesize that these genes increase the permeability of the intestinal barrier. This could alter the way that gluten enters the body, leading to a different response by the immune system. This new knowledge does not directly lead to a new treatment, but may well help direct the search for new treatments. We have shown that there is a shared genetic background between different intestinal disorders and that this can help in identifying new genes for these disorders. In the future it will be important to find the other genes involved in CD, the biological role of these genes, and how they interact. In that respect it will be necessary to increase the patient and control cohorts and to cooperate with other groups that have large cohorts for replication studies. We need to increase the information that we have on the individuals in the cohorts in order to perform stratification studies and to find subgroups of patients that might have specific disease variants which are absent in other patient groups.

Samenvatting

Coeliakie is een complexe, multifactoriële genetische aandoening die vóórkomt bij ongeveer 1% van de bevolking. Een genetische gevoeligheid voor coeliakie houdt voor patiënten in dat zij niet een normaal westers dieet kunnen volgen, aangezien dat gluteneiwitten bevat. Deze gluten komen voor in tarwe, gerst en rogge en zijn de belangrijkste omgevingsfactor voor de opwekking van coeliakie. Het immuunsysteem van coelakiepatiënten die gluten eten reageert alsof deze een bedreiging voor het lichaam zijn. De reactie van het immuunsysteem leidt tot een vernietiging van het oppervlak van de dunne darm. Dit oppervlak is onder normale omstandigheden zeer groot vanwege de vele microscopisch kleine uitstulpingen en kan normaal gesproken alle noodzakelijke voedingsstoffen opnemen uit het gegeten voedsel. Echter, bij coelakiepatiënten die gluten eten wordt het darmoppervlak sterk verkleind doordat de uitstulpingen verdwijnen. Hierdoor zijn zij niet in staat voldoende voedingsstoffen op te nemen. Patiënten vertonen daardoor aan een scala aan symptomen die samenhangen met ondervoeding. Lang werd verondersteld dat de symptomen van coeliakie uitsluitend werden veroorzaakt door schade aan de darmen, maar tegenwoordig is bekend dat ook andere organen worden aangetast. De enige behandeling voor coeliakiepatiënten is een levenslang glutenvrij dieet dat het darmoppervlak de kans geeft te herstellen.

Naast deze omgevingsfactor zijn ook genvariaties betrokken bij de ziekte. Hiertoe horen algemene variaties die voorkomen in de hele bevolking. Coeliakiepatiënten hebben óf een verkeerde combinatie óf teveel van deze varianten, wat bij hen lijdt tot ontwikkeling van de ziekte. De *HLA-DQA1* en *-DQB1* genen vormen samen een eiwit dat door het immuunsysteem gebruikt wordt om binnengedrongen bacteriën of virussen te herkennen en te presenteren aan andere cellen van het immuunsysteem die deze indringers vervolgens opruimen. Deze genen bevatten meerdere variaties, waardoor eiwitten gevormd worden die verschillende bacteriën en virussen kunnen presenteren. Sommige van deze varianten vormen de DQ2.5 en DQ8 eiwitten, welke een speciaal talent hebben: deze zijn in staat gluten te presenteren aan het immuunsysteem. Indien gluten op deze manier gepresenteerd worden zal het immuunsysteem reageren alsof gluten een bedreiging tegen het lichaam vormen. Vrijwel alle coeliakiepatiënten hebben de DQ2.5 of de DQ8 eiwitten en zullen ziek worden indien zij gluten eten. Echter, meer dan 25% van de gehele bevolking bezit ook één van deze varianten, zonder daardoor coeliakie te ontwikkelen. Blijkbaar is het drager zijn van de DQ2.5 of DQ8 variant niet voldoende om ziek te worden. Coeliakiepatiënten zijn waarschijnlijk dragers van andere genvarianten, die samen met DQ2.5 en DQ8 coeliakie veroorzaken. Het zoeken naar sommige van deze genen en varianten was het belangrijkste onderzoeksdoel van dit proefschrift. Het is echter waarschijnlijk dat er nog veel meer varianten gevonden moeten worden.

Hoofdstuk 1.1 geeft een overzicht van de klinische en genetische aspecten van coeliakie. De beschrijving van onze zoektocht naar genen is in dit proefschrift in drieën gedeeld (delen 2, 3 en 4). Elk deel beschrijft genen die betrokken zijn bij een andere locatie in de darm.

In deel 2 onderzoeken we twee enzymen die betrokken zijn bij de afbraak van gluten in de darmholte.

Gluten zijn moeilijk verteerbare eiwitten omdat de meeste enzymen die eiwitten uit voedsel voor vertering knippen in kleinere delen niet in staat zijn om gluten te verwerken. Wij dachten dat coeliakiepatiënten wellicht varianten hebben van de specifieke glutenverterende enzymen die niet efficiënt zijn in de verwerking van gluten. We hebben gezocht naar verschillen tussen patiënten en een gezonde controlegroep voor twee van deze enzymen, pyroglutamyl-peptidase I (PGPEP1) en prolyl endopeptidase (PREP) (zie hoofdstukken 2.1 en 2.2), maar wij vonden geen verschillen in de genen of in de functie van de gevormde eiwitten. Ondanks het feit dat wij geen concreet bewijs hebben kunnen vinden voor de betrokkenheid van deze genen in de gevoeligheid voor coeliakie moeten we hun rol ook niet onderschatten. Een van de enzymen, PREP, wordt ook aangetroffen in sommige bacteriën. Op dit moment wordt onderzocht of de toevoeging van dit enzym aan het dieet van een patiënt ertoe kan bijdragen dat gluten al voordat zij de dunne darm binnenkomen in zulke kleine delen geknipt zijn dat het immuunsysteem er niet langer op reageert.

In deel 3 beschrijven wij onze zoektocht naar genen die betrokken zijn bij de vergrote doorlaatbaarheid van de darmwand bij coeliakiepatiënten.

Nadat we gekeken hebben naar de processen in de holte van de dunne darm hebben we ook gezocht naar genen die werkzaam zijn op de celwand die de scheiding vormt tussen het lichaam en de darmholte. Onze aandacht was niet direct gericht op deze barrière, maar werd hier naartoe getrokken door de resultaten die gepresenteerd worden in hoofdstuk 3.1. We zochten in het bij coeliakie betrokken gebied op 19p13.1 (de 'linkage' regio genaamd MYO9B, het voormalige CELIAC4) naar een gen betrokken bij coeliakie. Gebruikmakend van een eerste cohort van patiënten en controles vonden we een associatie met het gen *myosine IXB* (*MYO9B*), waar we al enig bewijs voor hadden en welke ook direct onder de top van de linkage regio gelocaliseerd was. Met behulp van een tweede cohort konden we deze resultaten repliceren, waarbij een p-waarde van $2{,}1*10^{-6}$ gevonden werd voor de gecombineerde cohorten. De kansverhouding van de allelen (*allelic odds ratio,* OR) voor *MYO9B* was 1,56 en personen die homozygoot zijn voor het risico-allel hadden een 2,3 keer

grotere kans om coeliakie te ontwikkelen. *MYO9B* kan ongeveer 25% verklaren van de genetische achtergrond van coeliakie, wat opgeteld kan worden bij de 40% die verklaard wordt door de *HLA-DQA1* en *-DQB1* genen. Er is weinig bekend over de werking van *MYO9B*, maar wij stellen, gebaseerd op de literatuur en onze kennis van coeliakie, dat dit gen betrokken is bij de vergroting van de doorlaatbaarheid van de darmwand (zie ook hoofdstuk 1.2 en de discussie in deel 5 voor meer informatie over het *MYO9B* gen).

De vondst van dit gen vergrootte onze kennis over coeliakie in meerdere opzichten. *MYO9B* was het eerste gen dat gevonden werd sinds de ontdekking van de betrokkenheid van de *HLA* genen bij coeliakie (naast het zwakke bewijs voor het *CTLA4* gen). In het begin was het meeste onderzoek gericht op het zoeken naar genen met een functie die interessant kon zijn vanuit het oogpunt van de ziekte, maar waarbij een kans bestond op het over het hoofd zien van genen met onbekende functies of genen met functies de we nooit in overweging hadden genomen voor coeliakie. *MYO9B* werd gevonden door gebruikmaking van een genoombrede aanpak die geen rekening hield met functies. De identificatie van *MYO9B* leidde bij ons tot een breder denkkader en tot onze hypothese dat *MYO9B* betrokken is bij de vergroting van de doorlaatbaarheid van de darmwand. We zoeken nog steeds naar een bevestiging van de stelling dat *MYO9B* ook betrokken is bij coeliakie in andere populaties, zoals blijkt uit hoofdstukken 3.2 en 3.3. Met onze nieuwe kennis over *MYO9B* en complexe genetische aandoeningen in het algemeen, besloten we om te bezien of *MYO9B* ook betrokken was bij andere darmaandoeningen. Dit leidde ertoe dat *MYO9B* ook werd geïdentificeerd als een factor voor colitis ulcerosa (p-waarde = $1.9*10^{-6}$; OR = 1.2). Dit werd bevestigd in andere populaties en is beschreven in hoofdstuk 3.4. Hoe *MYO9B* betrokken is bij de vergroting van de doorlaatbaarheid van de darmwand is onbekend. Mogelijk heeft het invloed op de zonula occludens (tight junction). Op basis hiervan vroegen wij ons af of genen die de zonula occludens vormen of invloed hebben op de werking ervan betrokken zijn bij coeliakie en chronische inflammatoire darmziekten (CID; een verzamelnaam voor de ziekte van Crohn en colitis ulcerosa). We waren in staat om aan te tonen dat twee van deze genen, *MAGI2* en *PARD3*, een rol spelen in de gevoeligheid voor coeliakie en colitis ulcerosa. Dit wordt beschreven in hoofdstuk 3.5 (resultaat voor coeliakie voor *PARD3*: p-waarde = $3.2*10^{-5}$, OR 1.23 en voor *MAGI2*: p-waarde = 0.0018; OR 1.26). Daarmee hebben we nu genen gevonden die betrokken zijn bij lichaamsfuncties die we niet eerder in ogenschouw hadden genomen en waren we in staat om aan te tonen dat er een gedeelde genetische achtergrond bestaat voor coeliakie en colitis ulcerosa.

In deel 4 beschrijven we de zoektocht naar genen die betrokken zijn bij de ontstekingskant van coeliakie en naar een goede methode om HLA risicofactoren op te sporen

Sommige darmaandoeningen laten een overlap zien tussen linkage regio's en een overlap tussen de betrokken biologische netwerken. Wij besloten deze informatie te gebruiken in onze zoektocht naar de hierbij betrokken genen. Het immuunsysteem speelt een rol bij zowel coeliakie als CID, en deze ziekten delen enkele linkage regio's. We onderzochten het hele 250 kb gebied op chromosoom 5q31, waarvoor al eens een associatie met CID is aangetoond, en we kozen ook een aantal immuungerelateerde genen in de regio op chromosoom 19, maar we vonden geen overtuigende associatie (zie hoofdstuk 4.1). Echter, tegelijkertijd vonden we het *MYO9B* gen in dit gebied.

Als laatste besloten we om de diagnostische testmethodes voor de bekende HLA varianten te verbeteren. Deze genen zitten in een gebied met een groot aantal varianten, wat het moeilijk maakt om te bepalen welke varianten de betrokken persoon bezit en om screening hiervoor mogelijk en betaalbaar te maken (wat behulpzaam kan zijn voor het stellen van de diagnose coeliakie, of om coeliakie uit te sluiten). We wisten dat er andere, eenvoudiger aan te tonen varianten in het gebied lagen en dat sommige van deze varianten altijd samen overerven, bijvoorbeeld met DQ2.5. Wij zochten naar deze varianten en testten of een of meer van deze varianten precies konden voorspellen of een persoon drager was van DQ2.5. Dit hebben we ook gedaan voor DQ2.2, DQ7 en DQ8, welke alle varianten zijn van de *HLA-DQA1* en *-DQB1* genen en welke alleen of in combinatie gevoeligheid voor coeliakie veroorzaken. Gebruikmakend van deze methode vonden we dat slechts zes 1-basepaarvariaties (*single nucleotide polymorphisms*, SNP's) nodig zijn om de DQ2.2, DQ2.5, DQ7 en DQ8 risicovarianten te voorspellen die voorkomen bij meer dan 95 procent van de coeliakiepatiënten. Voor deze voorspellende benadering vonden we een gevoeligheid >0,968, een specificiteit >0,994 en een voorspellende waarde >0,940. Onze methode is eenvoudig en kostenefficiënt en kan daardoor worden gebruikt voor het screenen van familieleden van coeliakiepatiënten of zelfs van hele populaties.

Deel 5 bevat een discussie over wat we geleerd hebben van het werk dat voor dit proefschrift is verricht.

De discussie van dit proefschrift is gericht op drie delen: de zoektocht naar genen die betrokken zijn bij complexe genetische aandoeningen, de rol van *MYO9B* bij coeliakie en wat we geleerd hebben van ons geneticaonderzoek naar coeliakie. De discussie eindigt met enkele toekomstperspectieven voor onderzoek naar coeliakie en

meer algemeen naar andere complexe genetische aandoeningen.
De zoektocht naar genen die betrokken zijn bij coeliakie is moeizaam omdat we niet verwachten mutaties te vinden, welke eiwitten veranderen, maar genetische variaties die de werking van eiwitten wijzigen en die ook in de gewone populatie aangetroffen worden. Daarnaast verschillen coeliakiepatiënten van elkaar in het soort varianten waarvan zij drager zijn (genetische heterogeniteit). Het gebruik van gezamenlijke overerving van varianten kan ook helpen bij het opsporen van de varianten zonder dat er voor alle varianten in het genoom getest moet worden. Echter, gezamenlijke overerving kan verhullen welke genvarianten de ware variant zijn die voor de gevoeligheid verantwoordelijk is. Hierdoor weten we nog steeds niet welke variant van *MYO9B* de ware is. De varianten die worden gebruikt voor genetische testen verschillen tussen populaties in allelfrequentie, wat leidt tot fout-positieve of fout-negatieve resultaten als de geteste en de controlegroep niet goed op elkaar afgestemd zijn. De moeilijkheid van het repliceren van de associatie tussen *MYO9B* en coeliakie is ook door ons beschreven en we beredeneren waarom een combinatie van genetische heterogeniteit, het niet testen van de werkelijke gevoelige variant en het gebruik van te kleine cohorten oorzaak kan zijn van de negatieve uitslagen van de replicatiestudies.

Ons belangrijkste resultaat is het aantonen van betrokkenheid van *MYO9B* in de genetica van coeliakie en colitis ulcerosa. We toonden ook aan dat twee zonula occludensgenen, *MAGI2* en *PARD3*, betrokken zijn bij coeliakie en colitis ulcerosa. Onze hypothese is dat deze genen de doorlaatbaarheid van de darmwand vergroten. Hierdoor verandert mogelijk de route die gluten het lichaam binnenleidt, wat een andere respons van het immuunsysteem veroorzaakt. Deze nieuwe kennis leidt niet direct tot een nieuwe behandelmethodiek maar zou zeer behulpzaam kunnen zijn bij de zoektocht hiernaar. We hebben laten zien dat verschillende darmaandoeningen een genetische achtergrond delen en dat dit kan leiden tot de identificatie van nieuwe genen betrokken bij deze ziekten. Toekomstig onderzoek is noodzakelijk om de andere bij coeliakie betrokken genen te vinden en om de biologische rol van en de interactie tussen deze genen te achterhalen. Daarom is het van belang de cohorten van patiënten en controles te vergroten en om samen te werken met andere onderzoeksgroepen die beschikken over grote cohorten voor replicatiestudies. Als laatste is het aan te bevelen de informatie die we over individuen in de cohorten hebben uit te breiden, zodat stratificatiestudies mogelijk worden en waardoor subgroepen patiënten gevonden kunnen worden met specifieke varianten van de ziekte die niet voorkomen bij andere patiëntgroepen.

Publications

This thesis

Monsuur AJ, de Bakker PIW, Pinto D, Verduijn W, Crusius JBA, Wijmenga C. Effective detection of HLA risk alleles in celiac disease using tag SNPs. *Submitted.*

Wapenaar MC, Monsuur AJ, van Bodegraven AA, Weersma RK, Bevova MR, Linskens RK, Howdle P, Holmes G, Mulder CJ, Dijkstra G, van Heel DA, Wijmenga C. Association of tight junction genes PARD3 and MAGI2 with gluten-sensitive enteropathy and ulcerative colitis implies a common barrier defect. *Submitted.*

Monsuur AJ, Wijmenga C. Understanding the molecular basis of celiac disease: What genetic studies reveal. *Annals of Medicine 2006;38:578 - 591.*

Monsuur AJ*, van Bodegraven AA*, Curley CR*, Hunt KA*, Linskens RK, Onnie CM, Crusius JB, Annese V, Latiano A, Silverberg MS, Bitton A, Fisher SA, Steinhart AH, Forbes A, Sanderson J, Prescott NJ, Strachan DP, Playford RJ, Mathew CG, Wijmenga C, Daly MJ, Rioux JD, van Heel DA. Genetic variation in myosin IXB is associated with ulcerative colitis. *Gastroenterology 2006;131:1768-74.*

Amundsen SS, Monsuur AJ, Wapenaar MC, Lie BA, Ek J, Gudjonsdottir AH, Ascher H, Wijmenga C, Sollid LM. Association analysis of MYO9B gene polymorphisms with celiac disease in a Swedish/Norwegian cohort. *Hum Immunol 2006;67:341-5.*

Monsuur AJ*, Curley CR*, Wapenaar MC, Rioux JD, Wijmenga C. A functional candidate screen for coeliac disease genes. *Eur J Hum Genet 2006;14:1215-22. doi:10.1038/sj.ejhg.5201687*

Hunt KA, Monsuur AJ, McArdle WL, Kumar PJ, Travis SP, Walters JR, Jewell DP, Strachan DP, Playford RJ, Wijmenga C, van Heel DA. Lack of association of MYO9B genetic variants with coeliac disease in a British cohort. *Gut 2006;55:969-72.*

Monsuur AJ, Stepniak D, Diosdado B, Wapenaar MC, Mearin ML, Koning F, Wijmenga C. Genetic and functional analysis of pyroglutamyl-peptidase I in coeliac disease. *Eur J Gastroenterol Hepatol 2006;18:637-44.*

Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, Franke L, van't Slot R, van Belzen MJ, Lavrijsen IC, Diosdado B, Daly MJ, Mulder CJ, Mearin ML, Meijer JW, Meijer GA, van Oort E, Wapenaar MC, Koeleman BP, Wijmenga C. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. *Nat Genet 2005;37:1341-4.*

Monsuur AJ*, Diosdado B*, Stepniak DT, Franke L*, Wapenaar MC, Mearin ML, Koning F, Wijmenga C. No genetic association of the human prolyl endopeptidase gene in the Dutch celiac disease population. *Am J Physiol Gastrointest Liver Physiol 2005;289:G495-500.*

Others

Wapenaar MC, Monsuur AJ, Poell J, van 't Slot R, Meijer JW, Meijer GA, Mulder CJ, Mearin ML, Wijmenga C. The SPINK gene family and celiac disease susceptibility. *Immunogenetics 2007;59(5):349-57.*

Monsuur AJ*, Diosdado B*, Mearin ML, Mulder C, Wijmenga C. The downstream modulator of interferon-gamma, STAT1 is not genetically associated to the Dutch coeliac disease population. *Eur J Hum Genet 2006;14:1120-4.*

de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke X, Monsuur AJ, Whittaker P, Delgado M, Morrison J, Richardson A, Walsh EC, Gao X, Galver L, Hart J, Hafler DA, Pericak-Vance M, Todd JA, Daly MJ, Trowsdale J, Wijmenga C, Vyse TJ, Beck S, Murray SS, Carrington M, Gregory S, Deloukas P, Rioux JD. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet 2006;38:1166-72.*

Wapenaar MC, Monsuur AJ, Wijmenga C. Erfelijke factoren bij coeliakie. *Tijdschrift voor Kindergeneeskunde 2004;72:11-16.*

Bakker SC, van der Meulen EM, Buitelaar JK, Sandkuijl LA, Pauls DL, Monsuur AJ, van 't Slot R, Minderaa RB, Gunning WB, Pearson PL, Sinke RJ. A whole-genome scan in 164 Dutch sib pairs with attention-deficit/hyperactivity disorder: suggestive evidence for linkage on chromosomes 7p and 15q. *Am J Hum Genet 2003;72:1251-60.*

*Shared First authorship

Acknowledgements

Bedankt!

Nu is het zover. Het harde werken heeft iets moois opgeleverd.
Tijdens mijn studie wist ik niet goed wat voor baan ik wilde. De vakken vond ik leuk, maar ik werd er niet echt enthousiast van. AIO worden, dat leek me niets, na mijn studie zou ik wel iets anders vinden wat me leuk leek. Ik moest nog een laatste stage doen en koos deze meer omdat genetica me wel interessant en vooral rustig en saai leek. De hele dag PCRen kon toch niet moeilijk zijn. Na één maand stage lopen bij Steven Bakker ontdekte ik dat er zoveel meer achter zat en na drie maanden was ik helemaal om, genetica was interessant! Richard Sinke, die de co-promotor van Steven was, kwam naar mij toe om te zeggen dat hij mij graag als AIO wilde, maar dat hij geen project had. Hij had met Cisca Wijmenga gesproken en die had wel een project voor mij. Na twee keer met haar te hebben gepraat mocht ik van haar AIO worden op dit project en mijn gevoel zei dat het wel goed zat.
Steven, bedankt voor het begeleiden van mijn stage, zonder deze stage en jouw begeleiding zou ik geen AIO zijn geworden.
Richard, bedankt voor de begeleiding en de kans die ik via jou kreeg om AIO te worden!

Beste Cisca, ook al hadden we elkaar pas twee keer echt gesproken, had je blijkbaar wel vertrouwen in mij. Promoveren bij jou betekende hard werken, maar ook het krijgen van veel kansen. Ik wil je bedanken voor alle begeleiding en hulp. We hebben af en toe een pittig gesprek gehad, waardoor ik ben gaan zien hoe ik mijn werk beter kon aanpakken en jou beter op de hoogte kon houden en jij (bewust of onbewust) doorkreeg dat positieve communicatie naar mij toe veel oplevert. Ik heb hier veel geleerd en me kunnen ontwikkelen. De eerste Illuminaberekeningen staan nog vers in mijn geheugen. De halve groep stond om mijn tafel heen om te zien wat er uit kwam. Na een paar seconden was duidelijk dat we het gen voor coeliakie (wat Martine al op het spoor was gekomen) echt hadden gevonden. Het resulteerde in een Nature Genetics publicatie waar je trots op mag zijn en waarvan ik blij ben dat ik daar een grote rol in mocht spelen.

Beste Sasha, Marianna, Erica, Lude, Begoña, Eric, Alfons, Jan, Dalila en Martin. Vanuit verschillende oogpunten hebben we samen aan het onderzoek naar coeliakie gewerkt. Bedankt voor jullie hulp, gesprekken, samenwerking en gezelligheid. Erica en Begoña, jullie hebben mij allebei op een moeilijk moment een luisterend oor geboden. Bedankt daarvoor! Beste Dalila, een paar maanden heb je bijgedragen aan

het onderzoek naar coeliakie. Eén van de resultaten ervan is te zien in figuur 2 van de discussie. Bedankt voor die figuur en het werk!

Beste Karen, Eric, Ruben, Alfons en Frederieke. Veel van het onderzoek zou niets zijn geworden als jullie het niet algemeen ondersteunden. Karen, je hebt vaak last van mij gehad als een bestelling door ons erg belangrijk werd gevonden en jij tijd vrij moest maken, bedankt! Ruben, tijdens mijn stage werkten we samen aan het onderzoek van Steven, later heb je een belangrijke rol gespeeld door samen met Marianna de Illuminaproeven voor mij te doen. Nature Genetics was niet mogelijk geweest als jij je best niet had gedaan
Beste Albertien, Majiec, Clara, Anne, Gosia, Linda, Flip, Jelena, Ynte, Dineke, Bart van der Sluis, Harm, Bart Jungerius, Behrooz, Bobby, Roel, Charlotte, Carolien en de vele studenten die de laatste jaren op het lab waren (ik hoop zo dat ik niemand vergeet…). Ik heb een goede tijd gehad binnen deze groep, bedankt voor jullie hulp en gezelligheid daarbij. Behrooz, krijg ik 5 euro van jou voor mijn proefschrift?
Beste Martine, je hebt dit onderzoek een geweldige start gegeven. Hopelijk ben je tevreden over het vervolg van het onderzoek.

Beste Jackie, Je hebt wat werk aan mij gehad. Vele artikels heb je nagekeken en bij de 'Summary' had ik er duidelijk geen nachtje over geslapen. Ik leerde wel om van te voren aan te kondigen dat iets nagekeken moest worden, maar soms moest je iets doen als er haast bij was en je eigenlijk geen tijd had. Toch deed je het altijd. Bedankt voor alle hulp en correcties!
Beste Harry, al die coeliakie patiënten die de database in moesten, samen met de lastige HLA gegevens, het was een aardige klus. Zonder dat had ik mijn onderzoek niet kunnen doen. Beste Judith, bedankt voor alle secretariële hulp die ik van jou kreeg, leuk dat je bij onze groep zit. Beste Victorien, succes met verzamelen van coeliakiepatiënten en het onderzoek!
Beste Bart Crusius en Willem Verduijn, het HLA onderwerp is lastige kost, maar erg leuk als je het eenmaal door hebt. Dank voor de medewerking aan mijn speciale HLA project.
Dear Paul de Bakker, David van Heel, John Rioux, Christine Curley, Mark Daly and Silja Amundsen, Thanks for all the projects that we have done together! Paul, I was relieved to hear from you how easy my calculations could be, if I would use Tagger. Cisca and I thought it would take so much time, while in the end it only took seconds. You have been important for my Nature Genetics paper and my HLA paper. Nice that I could meet you when you visited Holland to see your family or do work related

stuff. David, thank you for being part of the reading committee and for all the times that you helped during my PhD period. Silja, nice to work with you and to meet you. Good luck with the last things for your PhD and with the care for your son.
Beste Chris en Ad, twee gastro-enterologen uit Amsterdam die een belangrijke rol speelden in het onderzoek. Jullie zijn beiden gedreven, een goede eigenschap!

Beste Ineke, negen maanden lang heb je hard gewerkt aan het onderzoek naar coeliakie. Je was een ideale stage student die de boel goed aanpakte. Bedankt voor al het werk en de gezelligheid. Gelukkig heeft stage lopen bij mij je niet bij de wetenschap weg doen lopen. ;-) Ik hoop dat je het naar je zin hebt op je AIO plek.

Het coeliakie consortium, mijn onderzoek hoorde bij het werk van het consortium. Bedankt voor de samenwerking en het luisteren naar mijn presentaties.
De patiëntenvereniging, leuk dat ik bij jullie vele praatjes mocht houden op patiëntendagen. Ik heb ervan genoten.

Beste familie, dank voor het familie zijn. Hanneke, ik ben benieuwd of jij over een paar jaar ook AIO wordt? Mama, bedankt voor de gezellige bezoekjes tijdens mijn zwangerschapsverlof (en voor de schone ramen e.d. toen mijn buik zo in de weg zat dat ik daar geen puf meer voor had). En bedankt voor die vele oppasuurtjes. Je weet wel gek te doen met je kleinzoon. Beste pa, je dochter heeft haar ouders verlaten, maar kan af en toe nog wel wat goede raad gebruiken. Die gaf je me ook wel als het nodig was, maar ik mocht het ook naast me neer leggen als ik het er niet mee eens was. Volgens jou moet er nog zoveel lagen informatie in het DNA zitten die we niet kennen. Ik zal de ontwikkelingen wel vanaf de zijlijn blijven volgen en je krijgt vast gelijk, net als zovaak al wilde ik dat liever niet altijd toegeven.

Beste schoonfamilie, bedankt voor de familie die ik er bij heb gekregen. Gezelligheid is wel te vinden bij jullie. Beste Thea, bedankt voor oppasuurtjes. Ik liet Allard met een gerust hart bij jou achter toen ik mijn onderzoek af moest maken.

Beste gevleugelde vrienden, waar kun je het beste je ei kwijt als het onderzoek weer eens lastig is. Van drie stellen en één single, zijn we in de loop van mijn onderzoeksjaren uitgegroeid naar vier stellen, twee baby's (die vriendjes moeten worden, of ze het willen of niet…) en een d.v. f.t. baby. Jullie vriendschap is belangrijk voor mij. Wanneer gaan we weer een weekendje weg?
Beste Mirjam en Sander, Geweldig om vrienden te hebben die twee straten verderop

wonen. Bedankt voor het oppassen, de gezelligheid, en de heerlijk ongecompliceerde gesprekken. Ik kan eigenlijk niet bij jullie langslopen zonder er minstens een paar uur te blijven hangen.
Beste overige vrienden, bedankt voor alle tijd en gezelligheid die ik met jullie meemaak. Met vrienden kun je de goede en verdrietige momenten van het leven meemaken en jullie zijn er echt voor mij geweest.

Lieve Jan en lieve Allard, ik heb twee geweldige mannen in mijn leven!
Jan, hoe heerlijk is het om in de jaren dat ik vaak gestrest en druk was rust te vinden bij jou. Jij weet dat positieve en bemoedigende woorden erg effectief zijn voor mij en hoe ontspannend rug en schoudermassages zijn. Bedankt daarvoor! In het laatste jaar voor mijn promotie was ik zwanger en mede dankzij jou heb ik dat jaar genoten van mijn zwangerschap en van de geboorte van onze zoon Allard. Gelukkig had ik ook op tijd ontdekt dat stressen nergens toe leid en heb ik wat dat betreft een relaxed laatste promotie jaar gehad. Bedankt voor je hulp bij het maken van de kaft, lekker samen gefrustreerd achter Photoshop zitten... Bedankt voor je liefde, ik kan je niet meer missen!
Allard, op jouw geboortedag op 8 november heb ik tussen de weeën door nog een aantal laatste berekeningen gedaan voor een artikel, en die ook naar Cisca gemaild, en gezegd dat ik de weken erna niets voor het werk zou kunnen doen. Later die dag heb je me aardig hard aan het werk gezet en mocht ik om 23:45 jou in de armen sluiten. Wat een opluchting en geluk. Je was een grote baby en beloonde ons na een paar weken al met heerlijke, relatief lange nachten. Ik geniet elke dag van jou met je mooie glimlach en je verandert constant. Elke keer als jij het moeilijk met jezelf hebt dan zien we een week later wat de reden was; je leerde op je zij rollen, op je buik rollen, gefrustreerd op je buik liggen maar niet weten hoe je terug moet, en dan uiteindelijk leerde je terug te rollen. Mama houdt van jou!

De twee maanden na mijn zwangerschapverlof kon ik gebruiken om mijn proefschrift af te maken en ik heb mijn deadline voor de leescommissie gehaald en zoals jullie kunnen zien, is dit boekje ook op tijd naar de drukker gegaan. Iedereen bedankt voor alle hulp, zorg en aandacht!

Curriculum Vitae

Alida Jentien (Alienke) Monsuur werd op 23 maart 1978 geboren in Eemnes. In 1997 behaalde zij haar VWO-diploma aan het Dr. F.H. de Bruijne Lyceum te Utrecht. Ze begon daarna aan de studie Medische Biologie aan de Universiteit Utrecht. Tijdens deze studie was ze lid van de Navigators Studentenvereniging Utrecht, waar ze deelnam aan verschillende commissies en voorzitter was van de slotweekend en almanakcommissie. Hier heeft ze ook haar man Jan Wijmenga ontmoet waarmee ze op 25 september 2003 in het huwelijk trad. Ze heeft tijdens haar studie twee wetenschappelijke stages gedaan bij de Universiteit Utrecht; een zesmaands stage bij de Divisie Immunologie & Hematologie in het UMC waar ze werkte aan het isoleren van multipele myeloma specifieke T cel klonen van leukemie patiënten en een negenmaands stage bij de Divisie Biomedische Genetica, waar ze onderzoek deed naar de erfelijke factoren betrokken bij ADHD. Op 16 december 2002 behaalde zij haar MSc in de Biomedische Wetenschappen. Zij begon op 1 december 2002 bij de Sectie Complexe Genetica van de Divisie Biomedische Genetica als AIO. Hier verrichte ze het door NWO gesubsidieerde onderzoek naar de genetica van coeliakie. Tijdens het laatste deel van haar promotie onderzoek beviel ze van een zoon, Allard. Aan het einde van de 4 jaar rondde ze haar promotie onderzoek af en beschreef de resultaten in dit proefschrift.