The puzzle of coeliac disease: pieces of the molecular pathogenesis

ISBN 90-393-4180-X

The work presented in this thesis was performed at the DBG-Department of Medical Genetics of the University Medical Center Utrecht, within the Research School of Developmental Biology. Financial support was provided by the Netherlands Organisation for Medical and Health Research (NWO) grant 902-22-094.

Financial support for the publication of this thesis by the DBG-Department of Medical Genetics of the UMC Utrecht, NWO, the University of Utrecht, the Section of Experimental Gastroenterology (SEG) of the Netherlands Society of Gastroenterology (NVGE), AstraZeneca B.V., Ferring Pharmaceuticals B.V., Nutricia Nederland B.V., Dr Ir van de Laar Foundation and the J.E. Jurriaanse Foundation is gratefully acknowledged.

The puzzle of coeliac disease: pieces of the molecular pathogenesis

De puzzel van coeliakie: stukken van de moleculaire pathogenese

(met een samenvatting in het Nederlands)

El puzle de la enfermedad celiaca: piezas de su patología molecular

(con el sumario en español)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de Rector Magnificus, Prof. dr. W.H. Gispen ingevolge het besluit van het College voor Promoties in het openbaar te verdedigen op dinsdag 21 maart 2006 des namiddags te 14:30 uur

door

María Begoña Diosdado Calvo
geboren op 30 januari 1976, te Valladolid (Spanje)

Promotor: Prof. dr. C. Wijmenga

Complex Genetics Section
DBG - Department of Medical Genetics
University Medical Center Utrecht
Utrecht University, the Netherlands

A mis padres Gonzalo y Begoña,
a mis hermanos Beatriz y Rodrigo
y a Eric.

Contents

Coeliac disease (CD) is a chronic inflammatory disorder with a multifactorial predisposition that affects the small intestine. The known causative factors responsible for this disease are the environmental factor gluten, the genes that encode for the HLA-DQ2/DQ8 molecules on chromosome 6, and the *MYO9B* gene on chromosome 19. Together these genes can explain approximately 55% of the familiar aggregation. Gluten is a dietary protein able to trigger specific structural changes in the mucosa of the small intestine with a distinct sequence of development. Based on the description of these histological modifications, it has been established that the disease progresses through three stages: the first stage, called Marsh I, in which gluten triggers the activation of an inflammatory response in the mucosa, and that is thought to lead first to crypt hyperplasia (Marsh II stage) and eventually to villous atrophy (Marsh III stage). In trying to understand the molecular mechanisms that drive these histological changes, much attention has been focused on the role of inflammation in the disease process. Our initial research was based on previous genetic knowledge, in particular the role of HLA-DQ2/DQ8. These studies have contributed to our understanding of how gluten is recognized by HLA-DQ2/DQ8 molecules and how it activates specific αβ CD4+ T-cells located in the lamina propria. Later on the influence of tissue transglutaminase on the recognition of gluten peptides, and the involvement of the innate immune response through IL-15 at the epithelial layer of coeliac mucosa was studied. These studies led to different molecular pathways being proposed to explain how villous atrophy occurs, i.e. increased oxidative stress, increased apoptotic activity, or lost balance of the metalloproteinases activity. However, the approaches used had revealed technical limitations and were not always focused on the primary processes that take place in the intestinal mucosa of the individuals at risk of coeliac disease when they were exposed to gluten. Hence, the molecular mechanisms responsible for the pathogenesis of coeliac disease were only partially elucidated. Driven by new technological developments, we employed microarray technology on biopsies of coeliac patients in different stages of the disease and on controls to assess the expression of thousands of genes simultaneously. This strategy allowed us to ascertain new pieces of the puzzle, revealing the molecular mechanisms and pathways underlying the initial inflammation processes and the further progression towards mucosal transformation in coeliac disease. This has enabled us to identify putative causative genes.

Outline of this thesis

The first part of this thesis describes the current knowledge on the pathogenesis of CD. Chapter 1 focuses on the immunological response that takes place in the intestinal mucosa of the individuals susceptible to CD when they sense gluten. This review chapter summarizes the genes and types of studies that have been examined on the adaptive and innate immune response and also considers the pathways involved in the induction of oral tolerance. It then assesses how the immune response might evoke the histological damage seen in CD or whether this is a non-specific defence mechanism. A second review in chapter 2 concentrates on the pathways proposed for mediating villous atrophy in CD. This paper describes the molecular mechanisms thought to be involved in the restructuring of the villous crypt units, and discusses their solidity in the context of current insights from functional, genetic and

genomic studies.

The second part of this thesis concentrates on the identification of the molecular pathways underlying the pathogenesis of CD. In the first set of experiments, we compared Marsh III stage (MIII) coeliac biopsies to normal control biopsies to identify the genes involved in the tissue remodelling processes. Secondly, we investigated the first molecular events that occur after gluten withdrawal by comparing the expression profiles of MIII CD patients ingesting and later refraining from gluten (chapter 3). In chapter 4, we describe gene expression studies on biopsies from coeliac patients in different stages towards normalization after refraining from gluten intake, and from control individuals. This experiment revealed which immunological processes are modulated by gluten in the intestine of CD patients, and enabled us to monitor the transcriptional activity of the genes involved during the histological normalization of the coeliac mucosa. The processes that emerged from the differentially expressed genes were further corroborated by immunohistochemistry experiments.

The last part of this thesis aimed to identify putative primary genes in CD by integrating genetic and microarray data. The genetic studies conducted on two of the putative primary genes identified are described in chapters 5 and 6.

This thesis describes some novel pathways and genes involved in the pathogenesis of CD and the genetic studies carried out on two candidate genes that were found during this Ph.D. project. In chapter 7 the results are discussed and looked at in the context of some preliminary data from functional studies on the pathways that drive villous atrophy. Our view on the puzzle of coeliac disease pathogenesis is presented.

Part I

GENERAL INTRODUCTION

Chapter 1

Molecular mechanisms of the adaptive, innate and regulatory immune responses in the intestinal mucosa of celiac disease patients

Expert Rev Mol Diagn. 2005 Sept;5(5):681-700

Molecular mechanisms of the adaptive, innate and regulatory immune responses in the intestinal mucosa of celiac disease patients

Begoña Diosdado and Cisca Wijmenga†

Celiac disease is a complex genetic disorder that affects the small intestine of genetically predisposed individuals when they ingest gluten, a dietary protein. Although several genome screens have been successful in identifying susceptibility loci in celiac disease, the only genetic contributors identified so far are the human leukocyte antigen (HLA)-DQ2/DQ8 molecules. One of the most important aspects in the pathogenesis of celiac disease is the activation of a T-helper 1 immune response, when the antigen-presenting cells that express HLA-DQ2/DQ8 molecules present the toxic gluten peptides to reactive $CD4^+$ T-cells. Recently, new insights into the activation of an innate immune response have also been described. It is generally accepted that the immune response triggers destruction of the mucosa in the small intestine of celiac disease patients. Hence, the activation of a detrimental immune response in the intestine of celiac disease patients appears to be key in the initiation and progression of the disease. This review summarizes the immunologic pathways that have been studied in celiac disease thus far, and will point to new potential candidate genes and pathways involved in the etiopathogenesis of celiac disease, which should lead to novel alternatives for diagnosis and treatment.

Expert Rev. Mol. Diagn. 5(5), 681–700 (2005)

CONTENTS

†Author for correspondence
University Medical Centre, Complex Genetics Section, Stratenum 2.117, Department of Biomedical Genetics, PO Box 85060, 3508 AB Utrecht, The Netherlands
Tel.: +31 30 253 8427
Fax: +31 30 253 8479
t.n.wijmenga@med.uu.nl

KEYWORDS:
adaptive immune response, celiac disease, gluten, HLA-DQ2/8, innate immune response, oral tolerance

What is celiac disease?

Celiac disease (CD) is a disorder of the small intestine that affects genetically predisposed individuals when they ingest gluten, a dietary protein. Gluten is a storage protein rich in glutamine residues and that is necessary for the germination of the grains of wheat, rye and barley [1]. It also confers elasticity to bread and pasta dough during cooking. The gluten protein can be divided into two protein fractions, the nonalcohol-soluble glutenins and the alcohol-soluble gliadins, which is the fraction that causes toxicity to CD individuals [2]. Gluten was not common in the human diet until approximately 10,000 years ago when people started to settle in one place and became farmers. Since then, grains have replaced the basic foods of our Paleolithic ancestors, who ate mainly fruits, game, nuts, roots and legumes, and has become the base of the current Western food pyramid [3]. This sudden change in lifestyle and diet has been hard for genetic constitutions to follow, so the existence of CD patients nowadays might be just one consequence of the change. CD is one of the most common alimentary disorders, with prevalence between 0.5–2% in different populations [4]. CD was previously considered to be mainly a childhood disease, but during the past few years, the disease is also being frequently diagnosed among the adult population; the age of onset now varies from the first months of life, to the 7–8th decades of life [5]. Its clinical presentation varies from the classical malabsorption symptoms with diarrhea, anemia, vitamin deficiency, abdominal distension or failure to thrive in children, to a less classic presentation characterized by nonspecific fatigue or repetitive

abortions. Some patients are atypical or even asymptomatic. Although subclinical and atypical forms of the disease are increasingly being diagnosed, many CD cases go undiagnosed: approximately only one out of seven patients is diagnosed [6]. Part of the symptomatology is explained by the specific changes that gluten evokes in the mucosa of the patient's small intestine: lymphocytosis, crypt hyperplasia and villous atrophy. These histologic changes are the hallmark of the disease and they need to be present to make a diagnosis of CD [7]. The diagnosis is made according to the European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) criteria. The procedure includes a serology test for positive antiendomysium (EMA) and tissue transglutaminase (tTG) antibodies, and a biopsy of the small intestine that shows the histologic features [8]. The sensitivity and specificity of the EMA and tTG antibodies exceeds 90% and they can be used to screen populations and discover atypical CD patients [9], although a costly and uncomfortable small intestine biopsy is always needed to confirm the CD diagnosis. Hence, there is an urgent need for noninvasive tests, based on genetic or molecular characteristics, for screening large populations at low cost. Thus far, the only treatment for CD consists of a life-long gluten-free diet (GFD), and good compliance will reverse the clinical presentation as well as the histologic changes. However, although the diet is effective, it is expensive, hard to follow and needs continuous monitoring by dieticians. A few alternatives to the GFD have already been proposed for CD patients [10,11], and once the molecular aspects of the disease are elucidated, specific individual treatments should become possible in the future.

CD is a multifactorial disorder and both environmental factors, such as gluten, and a multigenic predisposition are needed to explain the etiology of the disease. The fact that genetic factors play a major role is evident from twin and family studies. The concordance rate of CD in dizygotic twins is 20%, compared with 86% in monozygotic twins [12], while first-degree relatives have a 20–60-fold increased relative risk, even though CD does not show a Mendelian pattern of inheritance [13]. The only known genetic factors are the human leukocyte antigen (HLA)-DQα1 (DQA1*05) and HLA-DQβ1 (DQB1*02) genes, which encode for the HLA-DQ2 heterodimer and are localized in the major histocompatibility complex (MHC) region on chromosome 6p21.3 (the CELIAC 1 locus). The HLA-DQ locus has been significantly linked in all the genetic studies conducted, and in a meta- and pooled analysis of several European populations [14–19]. Despite the fact that these results indicate an essential role for the HLA region in the etiopathogenesis of CD, the HLA genes only explain approximately 40% of the genetic contribution to CD [15,20] and several genome screens have established significant linkage to other chromosomal regions (CELIAC 2, on chromosome 5q31-33, CELIAC 3, on chromosome 2q23-32 and CELIAC 4 on chromosome 19p13.1) [14–19]. In addition, many functional candidate gene approaches have focused mainly on genes involved in immunologic processes. Most of these studies have yielded negative results, apart from conflicting data on the cytotoxic T-lymphocyte-associated protein 4 (*CTLA4*) gene on the CELIAC 3 locus [19,21–35].

Nowadays, genomic studies employing techniques such as microarrays are commonly used to identify both causative disease genes as well as molecular pathways. To date, only a few genome-wide expression profiling studies have been conducted in CD [36,37].

General molecular immunologic mechanisms of the pathogenesis of celiac disease

The small intestine contains the largest mucosal surface of the body that is in direct contact with the exterior, and this makes it highly vulnerable to pathogens. Therefore, it has to maintain a strong line of defense by being able to initiate both adaptive and innate immune responses through homing specific populations of lymphocytes to the intestinal mucosa and the structures of the gut-associated lymphoid tissue (GALT). Besides maintaining this defense system against pathogens, the small intestine also has to be able to distinguish food nutrients and commensal flora, and to evoke an immunosuppressive response to them to induce oral tolerance. Consequently, a lack of activation against pathogens might lead to gastrointestinal infections, while a breakdown of oral tolerance could lead to activation of a deleterious immune response against the host gut, as seen in Crohn's disease and CD [38].

The molecular mechanisms implicated in the pathogenesis of CD have been partly elucidated. Although gluten is a dietary protein that evokes oral tolerance in most people, in some individuals carrying HLA-DQ2/8 molecules it initiates a detrimental immune response on reaching the mucosa of the small intestine. This immune response is mediated by HLA-DQ2/DQ8 antigen-presenting cells (APCs) that recognize certain gluten peptides and present them to $CD4^+$ T-cells that infiltrate the lamina propria [13]. Recently, it has been shown that gluten also activates an innate immune response mediated by lymphokine-activated killer cells [39–41]. It has been proposed that the gluten peptides that are able to activate this innate system differ from the ones recognized by the HLA-DQ2/8 molecules [41]. Although the precise mechanisms have not been elucidated, it is now generally accepted that this immune response somehow triggers proliferation of the cells of the crypts (crypt hyperplasia) and loss of the villi (villous atrophy).

As HLA-DQ2/8 can only explain approximately 40% of the genetic contribution in CD, several studies have been focused on components in these immunologic pathways as additional causative targets in genetic studies, or have tried to determine how they could be involved in the pathogenesis of the disease by functional studies.

This review aims to summarize the pathways and genes that might introduce disturbances into the gut immune system that could be causally related to CD or otherwise involved in the pathogenesis of the disease.

Gut immune system & the induction of oral tolerance

The main function of the gut is to absorb nutrients, water and electrolytes, and to produce hormones. The gut has an extensive surface (400 m^2) and is in direct contact with the exterior [42].

These two important characteristics mean that, for its function, the gut mucosa contain the body's largest immune system, with specific anatomical and functional features, the GALT. This specialized immune system consists of Peyer's patches and the mesenteric lymph nodes [43].

Peyer's patches are macroscopic aggregates that consist of B-cell follicle centers surrounded by plasma cells, small B-cells, dendritic cells (DCs) and intermingled T-cells. These follicles are separated from the lumen by the follicle-associated epithelium, a cubical epithelial layer containing the microfold cells, a specific cell type of the Peyer's patches that works as a transporter of particular agents. Mesenteric lymph nodes are another important constituent of the intestine's immune system. They are connected with Peyer's patches via afferent lymph vessels and to the systemic circulation via efferent lymph vessels [43]. Due to the anatomical location and cell composition, both have been studied in the context of induction of oral tolerance and priming of naive T-cells. These studies have shown that in animal models with ablation of the Peyer's patches, there is no alteration of oral tolerance [44], thus Peyer's patches do not appear to be crucial in acquiring oral tolerance.

Antigens can enter the lamina propria through different routes, but once they are internalized, the gut immune system has to decide whether those antigens will be recognized as pathogens and then activate an immune response to eliminate them, or as beneficial and then induce a physiologic immune response and consequently, oral tolerance.

The induction of oral tolerance is affected by several other factors, such as dose and nature of the antigen, administration route, gender, age and genetic background of the individual [45–47]. This physiologic immune response varies depending on whether it is induced by high (>20 mg) or low (100–1000 ng) doses of antigen. The exact underlying molecular mechanisms remain elusive, but circumstantial evidence shows that when it is induced by high doses of antigen, there is an induction of clonal deletion or anergy of T-cells. One possible mechanism driving this deletion might be through Fas–Fas ligand (FasL)-mediated apoptosis of the T-cells [48], whereas anergy might occur when there is no proper T-cell receptor signal amplification due to the ligation of the T-cell receptor, in the absence of costimulatory molecules such as CD80, CD28 or CD86, or other molecules that are as yet unknown [49]. On the other hand, it has been proposed that precipitation of this immune response by low doses activates regulatory T-cells that will work to suppress the immune response. This could be performed by three $CD4^+$ T-cell subtypes: T-helper (Th)3, T-regulatory 1 cells and $CD4^+$ $CD25^+$ T-regulatory cells. Th3 cells may modulate their action mainly through transforming growth factor (TGF)-β and/or interleukin (IL)-10 [50], and T-regulatory 1 cells through IL-10 [51]. These two molecules are known to show immunosuppressive activity by inhibiting the differentiation of naive T-cells into effector Th1 cells and the production of Th1 inflammatory cytokines (e.g., interferon [IFN]-γ) or TGF-α. In addition, TGF-β induces immunoglobulin A secretion [52]. The molecular mechanisms of action of $CD4^+$ $CD25^+$ T-regulatory cells are as yet unknown. Some evidence comes from animal models that have shown that these cells produce CTLA4, IL-10 and TGF-β upon ovalbumin feeding [53], and from studies that showed expression of the transcription factor forkhead box (FOX)P3. This transcription factor is thought to block Th1 and Th2 differentiation [54], and individuals with genetic deficiencies in FOXP3 develop the immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) [55].

Lack of oral tolerance in celiac disease patients

It has been suggested that CD patients might not be able to generate oral tolerance to gluten or that the tolerance is broken down [38]. It is therefore interesting to review which pathways are involved in the induction of oral tolerance in CD and how they contribute to the disease. In biopsies of CD patients, TGF-β has shown intense staining in the subepithelial region, $CD8^+$ and $CD4^+$ T-cells of lamina propria compared with controls [56,57], or no changes [58]. However, mRNA levels and their distribution show no differences between patients and controls [56,58–61], and no genetic association to the *TGF-β* gene was reported in the Finnish population [62]. IL-10 shows conflicting results in CD biopsies. While one study, using immunohistochemistry and *in situ* hybridization, showed more positive staining of IL-10 in the villous and crypts of CD patients compared with controls [60], another study showed no changes in protein staining [63]. mRNA levels also do not show agreement on *IL-10* expression in CD biopsies [59–61]. Besides its putative involvement in the pathogenesis of CD, two studies on this gene have shown no association in the Finnish and Italian populations [62,64], but a marginal association has been reported in the Irish population (TABLE 1) [65].

Th3, T-regulatory 1 cells and $CD4^+$ $CD25^+$ T-regulatory cells play a key role in limiting the immune response's activation and by inducing oral tolerance to beneficial food antigens, commensal bacterial and other advantageous antigens. In CD, one study has shown decreased numbers of $CD4^+$ $CD25^+$ T-regulatory cells in the mucosa of untreated CD patients [66]. A key molecule involved in this process is CTLA4 [47].

The *CTLA4* gene has been widely followed in several CD populations. *CTLA4* is an interesting candidate gene due to its location in the CELIAC 3 region on chromosome 2q23-32 and its function in the immune system [14,19]. The CTLA4 protein is expressed by T-cells. When it binds to CD80 or CD86 molecules expressed on APCs, there is a lack of costimulatory signal to the antigen-stimulated T-cell receptor (TCR), together with a negative immunosuppressive signal that is sent through the TCR [67]. Djilali-Saiah and colleagues performed the first genetic study that reported association to the *CTLA4* gene in CD; they demonstrated association to the +49 polymorphism in the first exon of this gene [23]. Some studies in different populations have replicated this finding, while others have not, or have shown association to other variants in linkage disequilibrium to the +49 polymorphism [19,21–33,35,68]. Lastly, there have been no functional studies that could help elucidate the genetic contribution of *CTLA4* in CD (TABLE 1).

Table 1. Studies of cytokines involved in the introduction of oral tolerance in celiac disease .

Study	Cytokine/ interleukin	Type of study conducted	Samples tested	Results	Remarks	Chromosomal location	Ref.
Lionetti *et al.* (1999)	TGF-β	Immunohisto-chemistry	Biopsies from 10 CD-UT and 10 controls	Intense staining in villous tip epithelium in controls, and patchy and weak in CD, lamina propria of CD positively stained		19q13.1	[56]
Perez-Machado *et al.* (2003)	TGF-β	Immunohisto-chemistry	Biopsies from 9 CD and 20 controls	Reduced epithelial expression of TGF-β in CD		19q13.1	[58]
Hansson *et al.* (2002)	TGF-β	Immunohisto-chemistry	Biopsies from 5 CD-UT, 3 CD-T, 6 CD-challenged and 9 controls	Expressed in lamina propria and higher in CD-UT than disease control individuals		19q13.1	[57]
Perez-Machado *et al.* (2003)	TGF-β	*In situ* hybridization	Biopsies from 9 CD and 20 controls	No changes		19q13.1	[58]
Lionetti *et al.* (1999)	TGF-β	Quantitative real-time PCR	Biopsies from 10 CD-UT and 10 controls	No significant changes		19q13.1	[56]
Forsberg *et al.* (2002)	TGF-β	Quantitative real-time PCR	Biopsies from 25 CD-UT, 22 CD-T, 24 challenged CD and 30 controls	No changes in IEL or lamina propria lymphocytes		19q13.1	[60]
Nilsen *et al.* (1998)	TGF-β	Semiquantitative real-time PCR	Biopsies from 7 CD-UT, 13 CD-T and 12 controls	No changes		19q13.1	[61]
Lahat *et al.* (1999)	TGF-β	Semiquantitative real-time PCR	Biopsies from 12 CD-UT, 3 CD-T and 10 controls	Higher levels in CD-UT than controls		19q13.1	[59]
Woolley *et al.* (2005)	TGF-β	Genetic association	106 CD families, Finnish	No significant genetic association		19q13.1	[62]
Hansson *et al.* (2002)	TGF-β3	Immunohisto-chemistry	Biopsies from 5 CD-UT, 3 CD-T, 6 CD-challenged and 9 controls	Expressed in lamina propria and epithelium. Higher in individuals with a normal mucosa than CD-UT		14q24	[57]
Forsberg *et al.* (2002)	IL-10	Immunohisto-chemistry	Biopsies from 6 CD-UT and 6 controls	IL-10 cells were located in villous and crypts of CD patients, not in controls		1q31-q32	[60]
Beckett *et al.* (1996)	IL-10	*In situ* hybridization/ immunohisto-chemistry	Biopsies from 11CD-UT, 10 CD-T and 9 controls	No changes for either methods		1q31-q32	[63]
Forsberg *et al.* (2002)	IL-10	Quantitative real-time PCR	Biopsies from 25 CD-UT, 22 CD-T, 24 challenged CD and 30 controls	Higher in challenged CD than CD-T or controls. Higher in CD-UT than controls		1q31-q32	[60]

CD: Celiac disease; CD-T: Celiac disease patient – treated; CD-UT: Celiac disease patient – untreated; CTLA4: Cytotoxic T-lymphocyte-associated protein; GFD: Gluten-free diet; IEL: Intraepithelial lymphocyte; TGF: Transforming growth factor.

Table 1. Studies of cytokines involved in the introduction of oral tolerance in celiac disease (cont.).

Study	Cytokine/ interleukin	Type of study conducted	Samples tested	Results	Remarks	Chromosomal location	Ref.
Nilsen *et al.* (1998)	IL-10	Semiquantitative real-time PCR	Biopsies from 7 CD-UT, 13 CD-T and 12 controls	Below detection levels in CD and control individuals		1q31-q32	[61]
Lahat *et al.* (1999)	IL-10	Semiquantitative real-time PCR	Biopsies from 12 CD-UT, 3 CD-T and 10 controls	No changes		1q31-q32	[59]
Woolley *et al.* (2005)	IL-10	Genetic association	106 CD families, Finnish	No significant genetic association	Markers tested: IL-10 -819, -1082 and -592	1q31-q32	[62]
Hahn-Zoric *et al.* (2003)	IL-10	Genetic association	93 CD patients and 103 controls, Swedish	Significant association ($p < 0.05$)	Marker tested: IL-10-1087		[65]
Cataldo *et al.* (2003)	IL-10	Genetic association	32 CD and 96 controls, Italian	No significant genetic association	Markers tested: IL-10-1082, -819 and -592	1q31-q32	[64]
Hunt *et al.* (2005)	CTLA4	Genetic association	340 CD and 705 controls	Variant CTLA4 +1822T ($p = 0.019$), variant CT60 G ($p = 0.047$), haplotype with all variants ($p = 0.00067$)	Tested variants CTLA4 -1722, -658, -318, +49, +1822T and CT60 G	CELIAC 3	[21]
Naluai *et al.* (2000)	CTLA4	Genetic association	107 CD, Swedish and Norwegian families	Associated with the variant +49 A ($p = 0.012$), D2S2392 ($p = 0.037$), D2S2214 ($p = 0.044$)	Tested markers D2S1391, D2S2392, D2S116, D2S2214, D2S2237 and variant +49 (A/G)	CELIAC 3	[22]
Djilali-Saiah *et al.* (1998)	CTLA4	Genetic association	101 CD and 130 controls, French	Associated with the variant +49 A ($p = 0.002$)	Tested variant +49 (A/G)	CELIAC 3	[23]
Popat *et al.* (2002)	CTLA4	Genetic association	116 CD families, Northern Europe	No significant genetic association to the variant +49 A		CELIAC 3	[25]
Clot *et al.* (1999)	CTLA4	Genetic association	192 CD families, Italian and Tunisian	No significant genetic association to the variant +49 A ($p = 0.15$)		CELIAC 3	[24]
Holopainen *et al.* (1999)	CTLA4	Genetic association	100 CD families, Finnish	Associated with markers D2S2392 ($p = 0.017$) and D2S116 ($p = 0.0001$)	Tested markers D2S1391, D2S2392, D2S116, D2S2214, D2S2189, D2S2237, variant +49 (A/G) and CTLA (AT) n	CELIAC 3	[34]

CD: Celiac disease; CD-T: Celiac disease patient – treated; CD-UT: Celiac disease patient – untreated; CTLA4: Cytotoxic T-lymphocyte-associated protein; GFD: Gluten-free diet; IEL: Intraepithelial lymphocyte; TGF: Transforming growth factor.

Table 1. Studies of cytokines involved in the introduction of oral tolerance in celiac disease (cont.).

Study	Cytokine/ interleukin	Type of study conducted	Samples tested	Results	Remarks	Chromosomal location	Ref.
King *et al.* (2003)	CTLA4	Genetic association	149 family trios and 100 unrelated controls, UK	No significant genetic association	Tested variants CT60 and MH30	CELIAC 3	[35]
King *et al.* (2002)	CTLA4	Genetic association	166 CD families, UK	Association to the marker D2S2214 (p = 0.007) and D2S1391 (p = 0.032)	Tested markers D2S116, D2S2392, D2S2214, D2S2237, D2S1391, CTLA4 (AT) n, variant +49 (A/G) and variant -318	CELIAC 3	[28]
Mora *et al.* (2003)	CTLA4	Genetic association	199 CD and 144 controls, Italian	Association with the variant +49 A (p = 0.03)	Tested variant +49 (A/G)	CELIAC 3	[33]
Popat *et al.* (2002)	CTLA4	Genetic association	62 CD families, Swedish	Association with the variant +49 A (p = 0.02)	Tested variant +49 (A/G)	CELIAC 3	[26]
Rioux *et al.* (2004)	CTLA4	Genetic association	146 CD and 104 controls, Finnish	No association with the variant +49 A	Tested variant +49 (A/G)	CELIAC 3	[19]
Martin-Pagola *et al.* (2003)	CTLA4	Genetic association	41 CD families, Basque	No association	Tested +49 (A/G) and CTLA (AT) n variants	CELIAC 3	[29]
van Belzen *et al.* (2004)	CTLA4	Genetic association	215 CD and 215 controls, Dutch	Borderline association variant CT60 A/G (p = 0.048), no association variant +49 A/G	Tested +49 (A/G) and +CT60 A/G variants	CELIAC 3	[30]
Amundsen *et al.* (2004)	CTLA4	Genetic association	225 CD families, Swedish and Norwegian families	Association to variants -1147 T (p = 0.02), CT61 A (p = 0.03)	Variants tested: MH30, -1147, +49, CT60, CT61, JO31, JO30 and JO27-1	CELIAC 3	[31]
Haimila *et al.* (2004)	CTLA4	Genetic association	106 families, Finnish	No significant genetic association	Tested 25 polymorphic markers	CELIAC 3	[32]

CD: Celiac disease; CD-T: Celiac disease patient – treated; CD-UT: Celiac disease patient – untreated; CTLA4: Cytotoxic T-lymphocyte-associated protein; GFD: Gluten-free diet; IEL: Intraepithelial lymphocyte; TGF: Transforming growth factor.

The immune response in the gut needs to be tightly regulated. Any distortion in the dialog between the commensal flora and the epithelial layer, microbial overgrowth, loss of integrity of the epithelial layer, or altered priming or antigen presentation can lead to the breakdown of oral tolerance, induction of food allergies, inflammation or infections. Although a great effort has been made to define whether key regulatory genes of oral tolerance contribute causally to CD, it is still not known through which pathways this occurs in CD or what their contribution is to the etiopathogenesis of this disease. It has been speculated that modifications of gluten by the tTG could provoke breakdown of oral tolerance, but in general, it can be assumed that any other mechanism that would lead to the presentation of gluten peptides to $CD4^+$ T-cells of the lamina propria, such as lack of digestion of gluten peptides in the brush border or increased permeability of the intestinal barrier, could lead to the breakdown of oral tolerance to gluten peptides in CD. Determining these pathways might open up new avenues of therapy for CD patients.

Adaptive immune response in celiac disease

The adaptive immune response is a complex defense system that is only present in vertebrates, and its main characteristic is its specificity. It has two branches, a humoral response mediated by B-cells that produce antibodies against antigens and microbials, and a cell-mediated immunity that responds to infected cells and operates through two cell types: cytotoxic T-cells that recognize their target cells through the T-cell receptor and

destroy it by inducing its apoptosis, and Th cells that produce a variety of cytokines upon interaction with macrophages, DCs or APCs. The Th cells are divided into Th1 or Th2 depending on the profile of the cytokines produced. Th2 cytokines include IL-4, -5 or -13 and, in turn, they induce B-cells to produce antibodies. IL-12 induces polarization of naive T-cells into Th1 T-cells that will produce cytokines such as tumor necrosis factor (TNF)-α, IL-2 that induces proliferation of $CD4^+$ T-cells, and IFN-γ that participates in activating specific pathways, such as the Janus tyrosine kinase (Jak)–signal transducers and activators of transcription (STAT) pathway [69].

In CD, the harmful activation of the adaptive immune response to gluten is one of the key factors in the pathogenesis of the disease. This activation starts when the gluten peptides that have reached the lamina propria are recognized by APCs, and are exclusively presented by the HLA-DQ2 and/or -DQ8 molecules to an αβ $CD4^+$ T-reactive repertoire [13]. Besides, it has also been proven that these gluten peptides can be deamidated by the tTG enzyme present in the lamina propria and brush border of the mucosa. The enzyme's activity on the gluten peptides leads to modification of glutamine residues into negatively charged glutamic acid residues at anchor positions, thereby enhancing the binding to HLA molecules and their presentation [70]. In addition, this modification can increase the variety of target peptides at the lamina propria, and consequently lowers the threshold at which a susceptible person develops the disease [71,72].

The presentation of gluten peptides to $CD4^+$ T-cells leads to a Th1 response [61]. The mediators of the polarization of naive T-cells into Th1 cells in CD appear to be T-bet [73], IL-18 [74] and IFN-α [75], as shown by their high expression levels in biopsies of CD patients compared with controls. Surprisingly, although IL-12 is, in general, an important player during the polarization of naive T-cells into Th1 cells, the low levels of the Th1 cytokine IL-12 and the lack of activation of the STAT-4 protein do not implicate this cytokine in Th1 polarization in CD [61,76]. This is in agreement with the genetic studies that also show no genetic association to the *IL12B* gene [61,77,78]. TNF-α is a proinflammatory cytokine of the Th1 immune response, and it plays multiple roles in the immune response by enhancing chemotaxis of cytokines to inflamed sites and producing damage to the epithelial layer. Immunohistochemistry and *in situ* hybridization studies have shown that it is expressed more highly in CD patients than controls and mainly in the lamina propria [57,79–82], although mRNA levels show conflicting results [59–61]. In addition, *TNF-α* is located in the HLA region, and since there is speculation of a second susceptibility gene within this region, a number of association studies have been carried out. Some studies have shown that there is a modest association to *TNF-α* in the CD population, while others have shown no association (TABLE 1) [62,83–87]. These conflicting results might reflect how difficult it is to carry out genetic studies on the HLA-DQ region in CD, since a correct stratification on the HLA DR3/DR7 haplotypes has to be taken into account (due to the high linkage disequilibrium between the *HLA-DQ2* and other interesting genes of the region). Overall, it can be concluded that although *TNF-α* might have a putative causative role in some populations, functional studies do not show clear evidence that this gene plays a causal role in CD pathogenesis (TABLE 2).

IFN-γ coordinates many aspects of the innate and adaptive immune responses and is the principal cytokine produced by the αβ $CD4^+$ reactive T-cells upon gluten activation. *IFN-γ* mRNA levels are consistently higher in CD patients than controls [59–61,73,88–90], and *in situ* hybridization [80,82,91,92] and immunohistochemical studies have demonstrated that its expression is restricted to the cells of the lamina propria [58,60,61,80,93]. Lastly, although Wapenaar and colleagues demonstrated that mRNA levels of *IFN-γ* remain elevated in the normalized mucosa of CD patients, no positive genetic association to this gene has been found to date, thus suggesting that *IFN-γ* might not be a causative gene in CD (TABLE 2) [62,89]. Furthermore, it is tempting to consider that IFN-γ might also have a role in the tissue destruction observed in CD by stimulating the extensive activation of the adaptive immune response in the lamina propria that precedes the tissue remodeling [94]. IFN-γ sends signals through the Jak–STAT pathway, while the binding of IFN-γ to its receptor provokes the phosphorylation of the tyrosine kinase Jak proteins (Jak1, Jak2, Jak3 and TYK2) and, subsequently, of STAT proteins (STAT-1–6). The STAT proteins are transcription factors that translocate to the nucleus where they interact with IFN regulatory factors (IRF1–8), thereby modulating the transcription of cytokines (such as RANTES and IL-6) and other transcription factors (T-bet) and growth factors [95]. This entire pathway has been proven to be activated in the $CD4^+$ T-cells and epithelial cells of CD biopsies [96]. In addition, the IRF-1 has also been found to be upregulated in CD compared with controls [97], although genetic studies have again failed to show association [77]. However, one of the explanations for these negative results might be that the studies were underpowered. Since CD is a complex genetic disorder, it is expected that multiple genes with a low relative risk contribute to the disease. Therefore, in order to assess a positive association from any of these genes in CD, a larger sample set than those used in previous studies will be needed. A second explanation might be the random selection of the tested markers and the lack of information on local linkage disequilibrium structures at the time of selection. Given the above, the true causative factor might have been overlooked.

Finally, in order to define the cytokine profile in CD, several groups have studied the presence of Th2 cytokines such as IL-6 and IL-4, and IL-5 in biopsies of CD individuals; however, all these studies show conflicting results.

IL-6 is generally accepted to be a proinflammatory cytokine and participates in the Th2 immune response by activating B-cells [70]. Immunohistochemistry studies have shown either higher staining of IL-6 in the lamina propria and epithelium of untreated CD compared with treated CD patients or controls [79], or only expression on enterocytes, but no change of expression in the lamina propria between treated and untreated patients and controls [98]. One *in situ* hybridization study also showed

Table 2. Studies of cytokines involved in the immune response in celiac disease .

Study	Cytokine/ interleukin	Type of study conducted	Samples tested	Results	Remarks	Chromosomal location	Ref.
Przemioslo *et al.* (1994)	TNF-α	Immunohisto-chemistry	Biopsies from 11 UT-CD, 9 CD-T and 11 controls	Elevated staining in CD-UT than CD-T or controls in lamina propia and epithelium		CELIAC1	[79]
Westerholm *et al.* (2002)	TNF-α	Immunohisto-chemistry	Biopsies from 9 potential CD, 8 CD and 8 controls	Higher expression in CD than potential CD or controls		CELIAC1	[80]
Hansson *et al.* (2002)	TNF-α	Immunohisto-chemistry	Biopsies from 5 CD-UT, 3 CD-T, 6 CD-challenged and 9 controls	Expressed in lamina propria and higher in CD than control disease individuals		CELIAC1	[57]
Kontakou *et al.* (1995)	TNF-α	*In situ* hybridization	Biopsies from 5 UT-CD, 5 CD-T and 5 controls	Elevated expression in CD-UT than CD-T or controls in lamina propria and between CD-UT and controls in epithelium		CELIAC1	[81]
Kontakou *et al.* (1995)	TNF-α	*In situ* hybridization	Biopsies from 4 challenged CD	Elevated expression in lamina propria after 4 h		CELIAC1	[82]
Forsberg *et al.* (2002)	TNF-α	Quantitative real-time PCR	Biopsies from 25 CD-UT, 22 CD-T, 24 challenged CD and 30 controls	Higher expression in controls than CD-UT		CELIAC1	[60]
Nilsen *et al.* (1998)	TNF-α	Semiquantitative real-time PCR	Biopsies from 7 CD-UT,13 CD-T and 12 controls	Below detection levels in CD and control individuals		CELIAC1	[61]
Lahat *et al.* (1999)	TNF-α	Semiquantitative real-time PCR	Biopsies from 12 CD-UT, 3 CD-T and 10 controls	Higher levels in CD-UT than controls		CELIAC1	[59]
De la Concha *et al.* (2000)	TNF-α	Genetic AFBAC	71 CD families, Spanish	Significant genetic association ($p = 0.04$)	Polymorphisms tested: TNF-α -376, -308 and -238	CELIAC1	[84]
Woolley *et al.* (2005)	TNF-α	Genetic AFBAC	106 CD families, Finnish	Significant genetic association	Tested TNF-α -308 polymorphism	CELIAC1	[62]
Garrote *et al.* (2002)	TNF-α	Genetic association	66 cases and 63 controls, Spanish	Significant genetic association ($p = 0.0028$)	Tested TNF-α -308 polymorphism	CELIAC1	[85]
Polvi *et al.* (1998)	TNF-α	Genetic association	68 cases and 114 controls, Finnish	No significant genetic association	TNF-α2 polymorphism	CELIAC1	[83]
Louka *et al.* (2003)	TNF-α	Genetic association	327 cases, Norwegian and Swedish	Significant genetic association to DR3-DQ2 haplotype ($p = 0.003$)	Tested TNF-α -308 polymorphism	CELIAC1	[86]
Westerholm *et al.* (2002)	IFN-γ	Immunohisto-chemistry	Biopsies from 9 potential CD, 8 CD and 8 controls	Higher expression in CD than potential CD or controls		12q14	[80]

AFBAC: Affected family-based controls; CD: Celiac disease; CD-T: Celiac disease patient – treated; CD-UT: Celiac disease patient – untreated; IEL: Intraepithelial lymphocyte; IFN: Interferon; TNF: Tumor necrosis factor.

Table 2. Studies of cytokines involved in the immune response in celiac disease (cont.).

Study	Cytokine/ interleukin	Type of study conducted	Samples tested	Results	Remarks	Chromosomal location	Ref.
Forsberg *et al.* (2002)	IFN-γ	Immunohisto-chemistry	Biopsies from 5 CD-UT and 6 controls	IFN-γ positive cells were mainly located in the villous and more frequently in CD than controls		12q14	[60]
Al-Dawoud *et al.* (1992)	IFN-γ	Immunohisto-chemistry	Biopsies from 21 CD-UT, 13 CD-T and 32 controls	Increased expression of lymphocytes of lamina propria		12q14	[93]
Nilsen *et al.* (1998)	IFN-γ	Immunohisto-chemistry	Biopsies from 7 CD-UT, 13 CD-T and 12 controls	Higher number of IFN-γ cells in lamina propria in CD-UT than controls		12q14	[61]
Perez-Machado *et al.* (2003)	IFN-γ	Immunohisto-chemistry	Biopsies from 9 CD and 20 controls	Increased number of positive cells in CD compared with controls		12q14	[58]
Kontakou *et al.* (1995)	IFN-γ	*In situ* hybridization	Biopsies from 4 CD-T	Elevated expression in lamina propria after 4 h		12q14	[82]
Westerholm *et al.* (2002)	IFN-γ	*In situ* hybridization	Biopsies from 9 potential CD, 8 CD and 8 controls	Signal higher in CD than controls, and located in lamina propria		12q14	[80]
Veres *et al.* (2003)	IFN-γ	*In situ* hybridization	Biopsies from 5 CD patients and 5 controls	Higher expression in lamina propria of CD than controls		12q14	[91]
Kontakou *et al.* (1994)	IFN-γ	*In situ* hybridization	Biopsies from 5 CD-UT, 5 CD-T and 5 controls	Increased expression of lymphocytes of lamina propria from CD-UT to controls or CD-T, but not between CD-T and controls.		12q14	[92]
Breese *et al.* (1994)	IFN-γ	Isolation of mononuclear cells	Biopsies from 7 CD, 8 controls and 2 non-CD with villous atrophy	Higher in non-CD patients with villous atrophy or CD than controls		12q14	[90]
Forsberg *et al.* (2002)	IFN-γ	Quantitative real-time PCR	Biopsies from 25 CD-UT, 22 CD-T, 24 challenged CD and 30 controls	Higher in challenged CD than CD-T or controls, and higher in CD-T/-UT than controls		12q14	[60]
Salvati *et al.* (2005)	IFN-γ	Quantitative real-time PCR	Biopsies from 26 CD and 21 controls	Higher expressed in CD than controls		12q14	[88]
Monteleone *et al.* (2004)	IFN-γ	Quantitative real-time PCR	Biopsies from 7 CD-UT and 7 controls	Higher expressed in CD than controls	Measured in extracted lamina propria lymphocytes	12q14	[73]
Wapenaar *et al.* (2004)	IFN-γ	Quantitative real-time PCR	Biopsies from 30 CD and 5 controls	Higher expressed in CD patients than controls		12q14	[89]

AFBAC: Affected family-based controls; CD: Celiac disease; CD-T: Celiac disease patient – treated; CD-UT: Celiac disease patient – untreated; IEL: Intraepithelial lymphocyte; IFN: Interferon; TNF: Tumor necrosis factor.

Table 2. Studies of cytokines involved in the immune response in celiac disease (cont.).

Study	Cytokine/ interleukin	Type of study conducted	Samples tested	Results	Remarks	Chromosomal location	Ref.
Nilsen *et al.* (1998)	IFN-γ	Semiquantitative real-time PCR	Biopsies from 7 CD-UT, 13 CD-T and 12 controls	Higher in CD-UT than CD-T or controls		12q14	[61]
Lahat *et al.* (1999)	IFN-γ	Semiquantitative real-time PCR	Biopsies from 12 CD-UT, 3CD-T and 10 controls	Higher levels in CD-UT than controls		12q14	[59]
Woolley *et al.* (2005)	IFN-γ	Genetic association	106 CD families, Finnish	No significant genetic association	Markers IFN-γ +874	12q14	[62]
Wapenaar *et al.* (2004)	IFN-γ	Genetic association	207 CD patients and 210 controls, Dutch	No significant genetic association		12q14	[89]
Przemioslo *et al.* (1994)	IL-6	Immunohisto-chemistry	Biopsies from 11 CD-UT, 9 CD-T and 11 controls	Elevated staining in CD-UT than CD-T or controls in lamina propria and epithelium		7p21	[79]
Jones *et al.* (1993)	IL-6	Immunohisto-chemistry	Biopsies from 3 CD-UT, 3 CD-T and 20 controls	Expressed on enterocytes, no changes in expression between CD and controls		7p21	[98]
Kontakou *et al.* (1995)	IL-6	*In situ* hybridization	Biopsies from 5 CD-UT, 5 CD-T and 5 controls	Elevated expression in CD-UT than CD-T or controls in lamina propria and between CD-UT and controls in epithelium		7p21	[81]
Kontakou *et al.* (1995)	IL-6	*In situ* hybridization	Biopsies from 4 challenged CD	Elevated expression in lamina propria after 4 h		7p21	[82]
Lahat *et al.* (1999)	IL-6	Semiquantitative real-time PCR	Biopsies from 12 CD-UT, 3 CD-T and 10 controls	Higher levels in CD-UT than controls		7p21	[59]
Nilsen *et al.* (1998)	IL-6	Semiquantitative real-time PCR	Biopsies from 7 CD-UT, 13 CD-T and 12 controls	Below detection levels in CD and control individuals		7p21	[61]
Woolley *et al.* (2005)	IL-6	Genetic association	106 CD families, Finnish	No significant genetic association		7p21	[62]
Westerholm *et al.* (2002)	IL-4	Immunohisto-chemistry	Biopsies from 9 potential CD, 8 CD and 8 controls	Higher expression in CD or potential CD than controls		CELIAC 2	[80]
Hansson *et al.* (2002)	IL-4	Immunohisto-chemistry	Biopsies from 5 CD-UT,3 CD-T, 6 CD-challenged and 9 controls	No changes		CELIAC 2	[57]
Nilsen *et al.* (1998)	IL-4	Immunohisto-chemistry	Biopsies from 6 CD-UT and 5 controls	No changes		CELIAC 2	[61]
Westerholm *et al.* (2002)	IL-4	*In situ* hybridization	Biopsies from 9 potential CD, 8 CD and 8 controls	Signal higher in controls than potential CD, equal signal between CD and controls, and signal located in lamina propria		CELIAC 2	[80]

AFBAC: Affected family-based controls; CD: Celiac disease; CD-T: Celiac disease patient – treated; CD-UT: Celiac disease patient – untreated; IEL: Intraepithelial lymphocyte; IFN: Interferon; TNF: Tumor necrosis factor.

Table 2. Studies of cytokines involved in the immune response in celiac disease (cont.).

Study	Cytokine/ interleukin	Type of study conducted	Samples tested	Results	Remarks	Chromosomal location	Ref.
Veres *et al.* (2003)	IL-4	*In situ* hybridization	Biopsies from 5 CD patients and 5 controls	Higher expressed in lamina propria of CD than controls		CELIAC 2	[91]
Beckett *et al.* (1996)	IL-4	*In situ* hybridization/ immunohisto-chemistry	Biopsies from 11CD-UT, 10 CD-T and 9 controls	No changes for either methods		CELIAC 2	[63]
Forsberg *et al.* (2002)	IL-4	Quantitative real-time PCR	Biopsies from 25 CD-UT, 22 CD-T, 24 challenged CD and 30 controls	No changes in IELs or lamina propria lymphocytes		CELIAC 2	[60]
Monteleone *et al.* (2004)	IL-4	Quantitative real-time PCR	Biopsies from 7 CD-UT and 7 controls	No changes		CELIAC 2	[73]
Nilsen *et al.* (1998)	IL-4	Semiquantitative real-time PCR	Biopsies from 7 CD-UT, 13 CD-T and 12 controls	Below detection levels in CD and control individuals		CELIAC 2	[61]
Lahat *et al.* (1999)	IL-4	Semiquantitative real-time PCR	Biopsies from 12 CD-UT, 3CD-T and 10 controls	No changes		CELIAC 2	[59]
Ryan *et al.* (2004)	IL-4	Genetic association	Biopsies from 409 CD patients and 355 controls, Irish	No significant genetic association		CELIAC 2	[99]
Desreumaux *et al.* (1998)	IL-5	Immunohisto-chemistry	Biopsies from 8 CD and 8 controls	Higher in eosinophils of CD than controls		CELIAC 2	[100]
Desreumaux *et al.* (1992)	IL-5	*In situ* hybridization	Biopsies from 4 CD-UT, 2 CD-T and 3 controls	Higher in eosinophils of CD-UT than CD-T or controls		CELIAC 2	[101]
Nilsen *et al.* (1998)	IL-5	Semiquantitative real-time PCR	Biopsies from 7 CD-UT, 13 CD-T and 12 controls	Below detection levels in CD and control individuals		CELIAC 2	[61]
Lahat *et al.* (1999)	IL-5	Semiquantitative real-time PCR	Biopsies from 12 CD-UT, 3CD-T and 10 controls	No changes		CELIAC 2	[59]
Ryan *et al.* (2004)	IL-5	Genetic association	Biopsies from 409 CD patients and 355 controls, Irish	No significant genetic association		CELIAC 2	[99]

AFBAC: Affected family-based controls; CD: Celiac disease; CD-T: Celiac disease patient – treated; CD-UT: Celiac disease patient – untreated; IEL: Intraepithelial lymphocyte; IFN: Interferon; TNF: Tumor necrosis factor.

that IL-6 is more highly expressed in the lamina propria of untreated CD patients than treated CD patients or controls [81], and furthermore, stimulation of CD biopsies also showed a positive signal in the lamina propria [82]. However, studies that quantify the mRNA levels are not in agreement, and while one study found more mRNA in biopsies from untreated CD patients than controls [59], another reported undetectable levels [61]. Lastly, no genetic association to the *IL-6* genes has been reported [62].

The main function of IL-4 is to differentiate cells into Th2 and B-cell activation. In the celiac mucosa, mRNA levels measured by quantitative and semiquantitative reverse transcriptase PCR showed no changes in expression for IL-4 between patients and controls [59–61,73], while two *in situ* hybridization studies showed higher expression in the lamina propria [80,91], and another reported no changes between patient and control biopsies [63]. Additionally, one immunohistochemistry study showed higher staining in CD

biopsies [80], while three reported no change [57,61,63], and a genetic association study in the Irish population showed no significant association [99].

IL-5 action mostly modulates the growth and differentiation of eosinophils. Consequently, *IL-5* was seen to be expressed in the eosinophils of CD patients [100,101], but its mRNA did not show changes between biopsies of patients and controls [59], nor could it even be measured [61], and a genetic association study reported negative association (TABLE 2) [99].

It is generally assumed that CD is a Th1-mediated disorder and that IFN-γ modulates its action. It appears that the downstream mediators (through which IFN-γ sends signals) have not been studied in detail in CD. Defining these patterns of cytokines could be crucial not only in characterizing the players in the Th1 immune response in CD, but also in helping elucidate how the immune response triggers tissue remodeling. Investigation of the presence of new cytokines and their role in the disease etiopathogenesis could also lead to discovering targets for therapy.

Innate immune response in the pathogenesis of celiac disease

The innate immune response is the first line of defense of the organism. The response is characterized by its immediate activation, and its lack of specificity and memory. In the gut, this first line of defense is the epithelial barrier. On the one hand, the epithelial layer functions as a compact physical barrier – with the microvilli and tight junctions preventing pathogens from passing through. On the other hand, the barrier can allow para- or intracellular passage of certain antigens (i.e., food antigens), and allow DCs to branch between the epithelial cells to sample antigens from the lumen [47]. The epithelial layer also has particular immunologic features mediated by Paneth cells, Globet cells, natural killer (NK), $CD8^+$ αβ, γδ T-cells and enterocytes that express pattern recognition receptors. Consequently, Paneth and Globet cells secrete antimicrobial peptides such as α- or β-defensins that breakdown Gram-positive and -negative bacteria, but also cathelicidins and mucins that bind to beneficial bacteria [102]. NK and $CD8^+$ αβ T-cells are cytotoxic cells, but the function of γδ T-cells is currently unknown. The enterocytes identify bacterial structures through the expression of pattern recognition receptors. These receptors are encoded in the germline and can be secreted or expressed on the cell surface, such as toll-like receptors (TLRs) and phagocytosis receptors. They recognize a variety of molecular structures known as pathogen-associated molecular patterns (PAMPs), which are mainly polysaccharides and PAMPs that contain mannose. There are 11 TLRs, and their engagement to PAMPs or other effector cytokines such as IL-1β or TNF can activate the nuclear factor (NF)-κB, mitogen-activated protein kinase (MAPK) and apoptotic pathways. The MAPK pathway is known to be activated by TLR-4, -5, -7 and -9 through different stimuli. Their binding to the TLR signals a cascade of phosphorylations leading to activation of the MAPK/extracellular signal-regulated kinase (ERK), Jun N-terminal kinase (JNK)/stress-activated protein kinase (SAPK), or p38 MAPK pathways that will modulate growth, differentiation, inflammation and apoptosis, depending on the stimuli [69].

Maiuri and colleagues proposed that activation of the innate immune response could be involved in the reaction to gluten. They showed Fas-positive enterocytes and migration of T-cells while blocking the immune response with CTLA4 immunoglobulin [103], and activation of macrophages and enterocytes before any sign of T-cell activation [104]. Interestingly, they reported a direct effect of a nonimmunodominant gliadin peptide on the small intestine that does not lead to activation of an αβ $CD4^+$ T-cell-mediated response, but to an earlier immune response mediated by IL-15 [41]. Additional evidence comes from a study on biopsies of HLA-DQ2 and -DQ8-negative CD patients that showed production of anti-EMA antibodies and intercellular adhesion molecule-1-positive lamina propria mononuclear cells, but no tissue remodeling. Since these patients are HLA-DQ2 and/or -DQ8 negative, the authors considered that other genetic factors, or an immune response of a different nature to the one mediated by the HLA molecules, might explain why these patients react to gluten. Interestingly, these results show that activation of this parallel immune response is not able to induce tissue remodeling [105].

Two recent studies have added more insights into the nature of this immune response. As proposed by Maiuri and colleagues, IL-15 appeared to be a major modulator of the activity of nonimmunodominant gluten peptides [41]. IL-15 is produced by epithelial cells, macrophages and DCs after gluten challenge, and although there is no difference in mRNA levels between CD patients and controls, its protein levels are elevated in CD individuals [106]. Meresse and colleagues [39] and Hue and colleagues [40] have shown that upregulation of IL-15 upon *in vitro* gliadin challenge induces overexpression of MHC class I polypeptide-related chain A (MICA) on the enterocytes of CD patients and drives the transformation of cytotoxic T-lymphocytes into lymphokine-activated killers and expression of the killer cell lectin-like receptor subfamily K, member 1 (NKG2D) receptor. On activation through MICA binding, NKG2D associates with death-associated protein (DAP)-10 and induces phosphorylation of the MAPK kinase pathway through ERK and JNK and lysis of the enterocytes. Despite the relevance of the *MICA* gene in the activation of the innate immune response, an expression study on mRNA in biopsies from CD patients and controls did not show changes in expression [107], and genetic studies have shown conflicting results due to the strong linkage disequilibrium of this gene to the HLA-DQ genes [62,87,108,109]. Interestingly, two studies show association of the allele MICA5.1 to atypical forms of CD in HLA-DR3/DR3- and -DQ2-negative patients (TABLE 3) [110,111].

In summary, it appears clear that there is an activation of the innate immune response that precedes or coincides with the adaptive response in the small bowel of CD patients. Nevertheless, further research needs to elucidate whether there is a mutated non-HLA gene mediating this response and whether this activation occurs in all CD patients; in which case, it can

Table 3. Studies on MICA in the innate immune response in celiac disease.

Study	Cytokine/ interleukin	Type of study conducted	Samples tested	Results	Remarks	Chromosomal location	Ref.
Woolley *et al.* (2005)	MICA	Genetic association	106 CD families, Finnish	No association was found independently of DQ genes		CELIAC 1	[62]
Fernandez *et al.* (2002)	MICA	Genetic association	157 CD and 305 controls, Spanish	No significant genetic association	Tested MCA5.1 allele	CELIAC 1	[108]
Rueda *et al.* (2003)	MICA	Genetic association	61 CD families, Spanish	Associated independently of DQ genes and clinical forms (p = 0.02)	Tested MCA5.1 allele	CELIAC 1	[109]
Lopez-Vazquez *et al.* (2002)	MICA	Genetic association	38 CD and 91 controls, Spanish	Associated with atypical forms of CD compared with typical forms (pc = 0.03), or controls (pc = 0.002)	Tested MCA5.1 allele	CELIAC 1	[110]
Lopez-Vazquez *et al.* (2005)	MICA	Genetic association	133 CD and 116 controls, Spanish	Associated with atypical forms of CD compared with the typical forms (pc = 0.00006)	Tested MCA5.1 allele	CELIAC 1	[111]
Louka *et al.* (2003)	MICA	Genetic association	944 families from Finland, Norway, Sweden, Italy and UK	No significant genetic association			[87]
Martin-Pagola *et al.* (2003)	MICA	Quantitative real-time PCR	21 CD-UT, 8 CD-T patients and 4 controls	No changes in expression levels		CELIAC 1	[107]

CD: Celiac disease; CD-T: Celiac disease patient – treated; CD-UT: Celiac disease patient – untreated; MICA: Major histocompatibility complex class I-related chain A.

be considered as essential in the immunopathogenesis of CD. In addition, determining which other cell types and pathways are involved, and which receptors recognize gluten will also help define the links between the adaptive and innate immune responses in the mucosa of CD patients.

Link between the immune response & histologic damage

CD has two aspects to its pathogenesis: the activation of the immune response and the development of specific histologic damage. Although it is generally assumed that the immune response leads to the tissue damage in CD, the link between the two remains an unknown factor.

The pathway that causes villous atrophy and crypt hyperplasia has not yet been elucidated, but different mechanisms have been proposed, including: an enhancement of apoptosis of the epithelial barrier; destruction of the extracellular matrix by metalloproteinases (MMPs); alterations in oxidative metabolism; and blocking of terminal differentiation of epithelial cells (reviewed in [112]). In order to define how the immune response might trigger tissue damage through any of these suggested mechanisms, the four different pathways proposed and their crosstalk to the immune system will be briefly discussed.

Villous disappearance in CD has been mainly attributed to an increased death of the cells of the villi, and apoptosis has therefore been studied in the biopsies of CD patients. The immunologic link between the apoptotic pathway and villous atrophy has been explained by two hypotheses. The first proposes that activated cytotoxic $CD8^+$ αβ intraepithelial T-cells that express FasL on their surface may interact with Fas-positive, stressed enterocytes and kill them. This hypothesis has been widely adopted and several other studies have indeed shown enhanced expression of Fas and FasL at the epithelial layer and increased apoptosis by the terminal deoxynucleotidyl transferase (TdT)-mediated dUTP nick-end labeling (TUNEL) technique [113–117]. The second hypothesis is that the innate immune response drives apoptosis of the enterocytes through the transformation of cytotoxic $CD8^+$ αβ T-cells into lymphokine-activated killers and activation of the MICA–NKG2D pathway [39,40].

A second mechanism involves MMPs, which are endopeptidases responsible for the maintenance of the extracellular matrix. This hypothesis proposes loss of the structure of the intestinal mucosa due to increased activity of the destructive properties of the MMPs versus their inhibitors (tissue inhibitors of metalloproteinases [TIMPs]), and consequent extracellular matrix destruction [118,119]. The link between the MMPs and the immune response comes from a fetal gut study, which showed that cytokines produced by activated T-cells of the lamina propria increased the production of MMPs and decreased the production of TIMPs [120,121].

A third mechanism consists of an accumulation of free radicals, due to increased free radical production or decreased reductive potential that can cause direct cell damage. In CD it has been shown that wheat gliadin can increase free radicals such as carbonyl groups, 4-hydroxy-2(E)-nonenal (4-HNE)

and protein sulfydryl groups [122]. This accumulation has two possible origins: first, it has been reported that gluten can induce production of free radicals and cell damage in cell lines [122,121]; second, decreased gluthatione peroxidase and reductase capacity, accumulation of stress markers and decreased antioxidants has been reported in CD biopsies [124–126]. Furthermore, gluten and the transcription factors of the immune response, NF-κB, STAT-1 and IRF-1, enhance the expression of nitric oxide synthase, an enzyme that produces the free radical nitric oxide [127]. Based on these data, it is evident that there is an altered oxidative stress in CD, but it is unclear how far this is directly responsible for the disappearance of villous or how important it is in mediating and maintaining the immune response by facilitating the contact of the APCs to the $CD4^+$ T-cells.

A fourth mechanism suggests that the cause of villous atrophy is a malfunction in the switch from proliferation to terminal differentiation, so that cells in the crypts might not be able to stop proliferating, and therefore they do not go on to differentiate and migrate out to reach the villous and maintain its structure [36]. This hypothesis is in line with previous ideas by Ferguson, who proposed that the primary lesion in CD is at the crypt and not at the villous. Support for this hypothesis comes from the observation that CD patients show hyperplasia at an earlier stage than villous atrophy. Moreover, it is tempting to speculate that released mediators of the immune response might alter the rates of proliferating and differentiated cells of the crypts. One of the key molecules of the immune response that also operates in the proliferation/differentiation pathways is TGF-β [128]. As explained earlier, TGF-β is an important molecule in oral tolerance established by Th3 cells, but it also signals through the Smad cascade to generate differentiation of epithelial cells [128]. Interestingly, this protein is expressed at the top of the villous in normal individuals, while CD patients only show a patchy staining at that location, which might imply a lack of differentiation.

An interesting, and potentially important, cell type that might play a pivotal role in CD are the stromal cells located underneath the epithelial layer and around the crypts. Stromal cells not only express growth factors such as keratinocyte growth factor, TGF-α, -β or epidermal growth factor necessary for maintaining the epithelial layer, but they are also immune response targets and thereby influence the expression of growth factors. For example, keratinocyte growth factor enhances proliferation, and its secretion can be increased by stromal cells under the influence of activated $CD4^+$ T-cells, such as in CD. In biopsies of CD patients, its mRNA expression is augmented compared with controls, although its expression was confined to cells of the lamina propria and underneath the atrophied villous, and was not seen in the pericryptal stromal cells [129].

Thus, if we assume that the immune response is in charge of directing the histologic damage in CD, it is more likely that more than one of these four mechanisms is responsible for the tissue remodeling. This could take place by activating some of the mechanisms in parallel and/or by their interaction with each other. However, although the immune response is thought to mediate the histologic changes in CD, we can also speculate that activation of a chronic immune response is not required to produce villous atrophy and crypt hyperplasia, but that this remodeling is a nonspecific defense reaction by the epithelial layer to an adverse situation. In this case, flattening of the mucosa and increased secretion of water and electrolytes could be seen as a nonspecific defense mechanism of the small bowel to an unwanted external antigen.

Expert commentary

During the last decade, the understanding of the immunopathogenesis of CD has increased tremendously and CD has become an example of how the causative genetic components of complex traits contribute to the pathogenesis of the disease. The genetic contribution of the MHC loci and the HLA-DQ2/DQ8 molecules to the pathogenesis of the disease has been well established [13].

Further genetic research has established that not all CD HLA-DQ2-positive individuals react to gluten in the same manner. Individuals with a DQ2.5 (DQA1*0501 and DQB1*0201) genotype have a higher risk of developing the disease than DQ2.2 (DQA1*0201 and DQB1*0202) carriers [71]. Therefore, since HLA-DQ2 typing may now be used as part of the diagnostic procedure, it is possible to perform more specific HLA-DQ2 typing for these other genotypes. This can enable classification of patients as high- or low-risk individuals and hence improve diagnoses, especially in atypical forms of the disease or in high-risk populations (e.g., first-degree relatives, patients with Down or Turner syndrome, diabetes mellitus or other autoimmune disorders).

One recently discovered aspect of the pathogenesis of CD is the activation of the innate immune response [39,40]. This interesting finding not only opens new lines of research, but also offers new perspectives as to how the pathogenesis of CD may function. Thus far, it is still not known whether this immune response is activated in all CD individuals or only in those carrying a certain genetic predisposition. Hence, one of the most intriguing questions is how gluten is recognized by the innate immune response and whether only specific gluten peptides initiate this response. From the pathogenesis point of view, it is important to define whether the activation of the innate response precedes that of the adaptive response, and to see how they modulate each other. Defining which other molecular pathways are involved in the innate immune response will also aid in understanding the extent that the innate response is contributing to the tissue destruction or influencing any of the mechanisms that drive crypt hyperplasia and villous atrophy. Finally, a better understanding of the pathogenesis will aid in finding new CD genes within the genetically linked regions.

This better understanding of the immune response not only raises new pointers to the pathogenesis, etiologic factors and diagnostic tests, but also possibilities for new therapies as alternatives to a life-long GFD, which is the only effective treatment currently available. These new alternatives include enzymotherapy to digest gluten into peptides too small to be presented

by HLA class II molecules [11], and genetic modification of the gluten-rich grains to reduce their toxicity for CD patients [10]. Another attractive possibility is a vaccine to suppress the adaptive immune response [130]. Technically, it should be possible to design a type of immunotherapy for CD, since it is known which HLA molecules are associated to the disease and how, which T-cell epitopes recognize gluten peptides and, very importantly, that the small intestine is easy to reach. The immune response could be targeted at different levels, thus, various possibilities can be envisaged. One level of action could be to silence the immune response by interfering with antigen presentation. This could be achieved by blocking the HLA-DQ2/DQ8 molecules so that they cannot recognize the toxic gluten peptides. Another possibility would be to change the gluten peptides by inhibiting the action of the enzyme tTG, or enhancing the digestion of gluten into very small pieces so they cannot be recognized by the HLA molecules, as suggested by Shan and colleagues [11]. Some of these approaches are on their way to becoming available as therapies to CD patients, such as an enhancement of the digestion of gliadin by adding a supplementary alimentary enzyme. Although inhibiting the activity of tTG might have important side effects, systemic administration of the enzyme in mice [131] or knockouts [132,133] have shown no deficiencies. However, these results could also imply that this treatment might not have a therapeutic effect. Lastly, there is HLA blocking, which might have unacceptable side effects and cannot be considered as a putative therapy for CD in the short term.

Another level of action would be to influence the precipitation/activation of the immune response by extinguishing the reactive T-cells of the intestinal mucosa. One way of achieving this is by inducing an immunosuppressive response in these patients at the site of inflammation [130]. This has already been carried out in other autoimmune disorders and is based on the principle of inducing regulatory T-cells against an antigen and migrating these cells to the target organ where they will suppress inflammation [134,135]. Experimental trials on animal models and even humans have shown that many factors can influence the successful outcome of this therapy, like the nature of the antigen, its concentration, method of administration and the stage of the disease process when the immunosupression is induced [47].

The authors consider this to be an interesting avenue for future research into alternative therapies for CD. Nonetheless, it is still necessary to define how CD patients lose their tolerance to gluten and the mechanisms through which this occurs. It is also important to investigate whether their genetic background (e.g., HLA and other CD genes), the stage of the disease or adjuvants of the Th1 immune response will affect potential new therapies for CD patients.

Five-year view

At present, one of the major questions in CD is which other non-HLA genes predispose to CD. The answer to this question has direct repercussions for the pathogenesis of the disease since it will allow researchers to answer key questions such as: how does gluten enter the intestinal barrier, what is the role of the innate immune response in CD, how does the immune response trigger tissue remodeling, and which molecular pathways does the immune response work through? In 5 years' time, it is expected that many of the other non-HLA genes will have been identified through genetic association studies performed

Key issues

- Celiac disease (CD) is a disorder of the small intestine that affects genetically predisposed individuals when they ingest gluten, a dietary protein.
- The human leukocyte antigen (HLA)-DQ2/DQ8 genes on chromosome 6p21 account for approximately 40% of the genetic contribution to CD. Although significant linkage has been reported to other regions in the human genome, no other genes have been identified as yet. Defining the molecular pathogenesis of the disease can help in elucidating other causative genes and will lead to new diagnostic and therapeutic approaches.
- In CD there might be a breakdown of oral tolerance to gluten that leads to the activation of a deleterious immune response against the host gut. No molecular pathways in oral tolerance that contribute to the pathogenesis of CD, or causative genes in these pathways, have been identified.
- Gluten activates a T-helper (Th)1 adaptive immune response mediated by the HLA-DQ2/DQ8 molecules that present gluten to $CD4^+$ α/β T-cells in the gut. Although it is known that interferon-γ is a key player in this response, further research is needed to find other immune mediators.
- It has recently been demonstrated that the innate immune response is also activated in CD patients. A large research effort should be made to find out whether a genetic predisposition for its activation is needed and to better characterize this response in CD.
- Thus far, none of the genes studied from immunologic pathways has shown positive genetic association in CD, but they may well be potential therapeutic targets.
- Although a better understanding of the molecular pathways that regulate CD has not yet indicated new causative genes, it has led to new diagnostic and therapeutic possibilities, such as molecular diagnostic tools or enzymotherapy.

on linkage regions, or through genome-wide association studies. More recently, it has become feasible to identify genes for complex disorders such as CD using well-characterized patient materials and improved technologies. These genes will enable a better understanding of the genetics of complex disorders, define the differences and similarities among CD populations, determine why not all HLA-DQ2 individuals develop CD, and why there are differences in clinical presentation and response to the treatment. Use of genomic and proteomic (i.e., gene and protein expression) technologies is expected to lead to the development of biomarkers that can be used to study the progression of the disease, and possibly also for diagnosis. Greater knowledge about the disease pathogenesis will have a direct impact on patients' daily care. Within 5 years, the authors expect to have a firm basis for developing better and more specific molecular diagnostic and prognostic tools, alternative treatments to the GFD, personalized treatment (pharmacogenomics and pharmacogenetics), and maybe even gene therapy with genetically engineered lymphocytes.

Acknowledgements

The authors gratefully acknowledge Jackie Senior and other staff for critical reading of the manuscript. The authors thank The Netherlands Organization for Medical and Health Research (NWO) grant 902-22-094, and the Celiac Disease Consortium (CDC), an innovative cluster approved by the Netherlands Genomics Initiative and partially funded by a Dutch government grant (BSIK03009), for financial support.

References

Papers of special note have been highlighted as:
- • of interest
- •• of considerable interest

1 Shewry PR, Halford NG. Cereal seed storage proteins: structures, properties and role in grain utilization. *J. Exp. Bot.* 53, 947–958 (2002).

2 Wieser H. Relation between gliadin structure and coeliac toxicity. *Acta Paediatr. Suppl.* 412, 3–9 (1996).

3 Eaton SB, Eaton SB III,Konner MJ. Paleolithic nutrition revisited: a twelve-year retrospective on its nature and implications. *Eur. J. Clin. Nutr.* 51, 207–216 (1997).

4 Rewers M. Epidemiology of celiac disease: what are the prevalence, incidence, and progression of celiac disease? *Gastroenterology* 128, S47–S51 (2005).

5 Dube C, Rostom A, Sy R *et al.* The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. *Gastroenterology* 128, S57–S67 (2005).

6 Fasano A, Catassi C. Current approaches to diagnosis and treatment of celiac disease: an evolving spectrum. *Gastroenterology* 120, 636–651 (2001).

7 Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 102, 330–354 (1992).

•• **Landmark paper about the histologic characteristics of celiac disease (CD) biopsies and classification of different stages of the disease.**

8 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).

9 Rostom A, Dube C, Cranney A *et al.* The diagnostic accuracy of serologic tests for celiac disease: a systematic review. *Gastroenterology* 128, S38–S46 (2005).

10 Vader LW, Stepniak DT, Bunnik EM *et al.* Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. *Gastroenterology* 125, 1105–1113 (2003).

11 Shan L, Molberg O, Parrot I *et al.* Structural basis for gluten intolerance in celiac sprue. *Science* 297, 2275–2279 (2002).

•• **First paper to show detoxification of a gluten peptide by a bacterial prolyl endopeptidase from *Flavobacterium meningosepticum*, and propose enzymotherapy as an alternative treatment to the gluten-free diet in CD.**

12 Greco L, Romino R, Coto I *et al.* The first large population based twin study of coeliac disease. *Gut* 50, 624–628 (2002).

13 Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. *Nature Rev. Immunol.* 2, 647–655 (2002).

14 Babron MC, Nilsson S, Adamovic S *et al.* Meta and pooled analysis of European coeliac disease data. *Eur. J. Hum. Genet.* 11, 828–834 (2003).

• **First meta- and pooled analysis performed on genome-wide screens carried out in four different European CD populations.**

15 Van Belzen MJ, Meijer JW, Sandkuijl LA *et al.* A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* 125, 1032–1041 (2003).

16 Greco L, Babron MC, Corazza GR *et al.* Existence of a genetic risk factor on chromosome 5q in Italian coeliac disease families. *Ann. Hum. Genet.* 65, 35–41 (2001).

17 Greco L, Corazza G, Babron MC *et al.* Genome search in celiac disease. *Am. J. Hum. Genet.* 62, 669–675 (1998).

18 Percopo S, Babron MC, Whalen M *et al.* Saturation of the 5q31–q33 candidate region for coeliac disease. *Ann. Hum. Genet.* 67, 265–268 (2003).

19 Rioux JD, Karinen H, Kocher K *et al.* Genomewide search and association studies in a Finnish celiac disease population: Identification of a novel locus and replication of the HLA and CTLA4 loci. *Am. J. Hum. Genet.* 130, 345–350 (2004).

20 Bevan S, Popat S, Braegger CP *et al.* Contribution of the MHC region to the familial risk of coeliac disease. *J. Med. Genet.* 36, 687–690 (1999).

21 Hunt KA, McGovern DP, Kumar PJ *et al.* A common CTLA4 haplotype associated with coeliac disease. *Eur. J. Hum. Genet.* 13, 440–444 (2005).

22 Naluai AT, Nilsson S, Samuelsson L *et al.* The CTLA4/CD28 gene region on chromosome 2q33 confers susceptibility to celiac disease in a way possibly distinct from that of Type 1 diabetes and other chronic inflammatory disorders. *Tissue Antigens* 56, 350–355 (2000).

23 Djilali-Saiah I, Schmitz J, Harfouch-Hammoud E *et al.* CTLA-4 gene polymorphism is associated with predisposition to coeliac disease. *Gut* 43, 187–189 (1998).

24 Clot F, Fulchignoni-Lataud MC, Renoux C *et al.* Linkage and association study of the CTLA-4 region in coeliac disease for Italian and Tunisian populations. *Tissue Antigens* 54, 527–530 (1999).

25 Popat S, Hearle N, Hogberg L *et al.* Variation in the CTLA4/CD28 gene region confers an increased risk of coeliac disease. *Ann. Hum. Genet.* 66, 125–137 (2002).

26 Popat S, Hearle N, Wixey J *et al.* Analysis of the CTLA4 gene in Swedish coeliac disease patients. *Scand. J. Gastroenterol.* 37, 28–31 (2002).

27 Holopainen P, Naluai AT, Moodie S *et al.* Candidate gene region 2q33 in European families with coeliac disease. *Tissue Antigens* 63, 212–222 (2004).

28 King AL, Moodie SJ, Fraser JS *et al.* CTLA-4/CD28 gene region is associated with genetic susceptibility to coeliac disease in UK families. *J. Med. Genet.* 39, 51–54 (2002).

29 Martin-Pagola A, Perez de Nanclares G, Vitoria JC *et al.* No association of CTLA4 gene with celiac disease in the Basque population. *J. Pediatr. Gastroenterol. Nutr.* 37, 142–145 (2003).

30 van Belzen MJ, Mulder CJ, Zhernakova A *et al.* CTLA4 +49 A/G and CT60 polymorphisms in Dutch coeliac disease patients. *Eur. J. Hum. Genet.* 12, 782–785 (2004).

31 Amundsen SS, Naluai AT, Ascher H *et al.* Genetic analysis of the CD28/CTLA4/ICOS (CELIAC3) region in coeliac disease. *Tissue Antigens* 64, 593–599 (2004).

32 Haimila K, Smedberg T, Mustalahti K *et al.* Genetic association of coeliac disease susceptibility to polymorphisms in the ICOS gene on chromosome 2q33. *Genes Immun.* 5, 85–92 (2004).

33 Mora B, Bonamico M, Indovina P *et al.* CTLA-4 +49 A/G dimorphism in Italian patients with celiac disease. *Hum. Immunol.* 64, 297–301 (2003).

34 Holopainen P, Arvas M, Sistonen P *et al.* CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to celiac disease. A linkage and family-based association study. *Tissue Antigens* 53, 470–475 (1999).

35 King AL, Moodie SJ, Fraser JS *et al.* Coeliac disease: investigation of proposed causal variants in the CTLA4 gene region. *Eur. J. Immunogenet.* 30, 427–432 (2003).

36 Diosdado B, Wapenaar MC, Franke L *et al.* A microarray screen for novel candidate genes in coeliac disease pathogenesis. *Gut* 53, 944–951 (2004).

37 Juuti-Uusitalo K, Maki M, Kaukinen K *et al.* cDNA microarray analysis of gene expression in coeliac disease jejunal biopsy samples. *J. Autoimmun.* 22, 249–265 (2004).

38 Sollid LM. Molecular basis of celiac disease. *Ann. Rev. Immunol.* 18, 53–81 (2000).

39 Meresse B, Chen Z, Ciszewski C *et al.* Co-ordinated induction by IL15 of a TCR-independent NKG2D signaling pathway converts CTL into lymphokine-activated killer cells in celiac disease. *Immunity* 21, 357–366 (2004).

• **Shows that interleukin (IL)-15 directs the innate immune response and the death of the enterocytes in CD through the expression of MHC class I polypeptide-related chain A (MICA) and killer cell lectin-like receptor subfamily K, member 1 (NKG2D) molecules at the epithelial layer.**

40 Hue S, Mention JJ, Monteiro RC *et al.* A direct role for NKG2D/MICA interaction in villous atrophy during celiac disease. *Immunity* 21, 367–377 (2004).

• **Shows that IL-15 directs the innate immune response and the death of the enterocytes in CD through the expression of MICA and NKG2D molecules at the epithelial layer.**

41 Maiuri L, Ciacci C, Ricciardelli I *et al.* Association between innate response to gliadin and activation of pathogenic T cells in coeliac disease. *Lancet* 362, 30–37 (2003).

•• **First identified nontoxic gluten peptide that is able to activate the innate immune response with IL-15 as a major mediator of the innate immune response.**

42 MacDonald TT. The mucosal immune system. *Parasite Immunol.* 25, 235–246 (2003).

43 Mowat AM. Anatomical basis of tolerance and immunity to intestinal antigens. *Nature Rev. Immunol.* 3, 331–341 (2003).

•• **Landmark paper about the anatomical structure of the gut immune system and how it functions in order to maintain correct homeostasis.**

44 Spahn TW, Fontana A, Faria AM *et al.* Induction of oral tolerance to cellular immune responses in the absence of Peyer's patches. *Eur. J. Immunol.* 31, 1278–1287 (2001).

45 Chehade M, Mayer L. Oral tolerance and its relation to food hypersensitivities. *J. Allergy Clin. Immunol.* 115, 3–12 (2005).

46 Strobel S, Mowat AM. Immune responses to dietary antigens: oral tolerance. *Immunol. Today.* 19, 173–181 (1998).

47 Mayer L, Shao L. Therapeutic potential of oral tolerance. *Nature Rev. Immunol.* 4, 407–419 (2004).

•• **Reviews how oral tolerance is generated at the mucosa of the small bowel and how it has been used as treatment in human disorders.**

48 Chen Y, Inobe J, Marks R *et al.* Peripheral deletion of antigen-reactive T cells in oral tolerance. *Nature* 376, 177–180 (1995).

49 Appleman LJ, Boussiotis VA. T cell anergy and costimulation. *Immunol. Rev.* 192, 161–180 (2003).

50 Weiner HL. Induction and mechanism of action of transforming growth factor-β-secreting Th3 regulatory cells. *Immunol. Rev.* 182, 207–214 (2001).

51 Groux H, O'Garra A, Bigler M *et al.* A $CD4^+$ T-cell subset inhibits antigen-specific T-cell responses and prevents colitis. *Nature* 389, 737–742 (1997).

52 Kim PH, Kagnoff MF. Transforming growth factor-β 1 is a costimulator for IgA production. *J. Immunol.* 144, 3411–3416 (1990).

53 Zhang X, Izikson L, Liu L, Weiner HL. Activation of $CD25^+CD4^+$ regulatory T cells by oral antigen administration. *J. Immunol.* 167, 4245–4253 (2001).

54 Ostroukhova M, Seguin-Devaux C, Oriss TB *et al.* Tolerance induced by inhaled antigen involves $CD4(^+)$ T cells expressing membrane-bound TGF-β and FOXP3. *J. Clin. Invest.* 114, 28–38 (2004).

55 Bennett CL, Christie J, Ramsdell F *et al.* The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nature Genet.* 27, 20–21 (2001).

56 Lionetti P, Pazzaglia A, Moriondo M *et al.* Differing patterns of transforming growth factor-β expression in normal intestinal mucosa and in active celiac disease. *J. Pediatr. Gastroenterol. Nutr.* 29, 308–313 (1999).

57 Hansson T, Ulfgren AK, Lindroos E *et al.* Transforming growth factor-β (TGF-β) and tissue transglutaminase expression in the small intestine in children with coeliac disease. *Scand. J. Immunol.* 56, 530–537 (2002).

58 Perez-Machado MA, Ashwood P, Thomson MA *et al.* Reduced transforming growth factor-β1-producing T cells in the duodenal mucosa of children with food allergy. *Eur. J. Immunol.* 33, 2307–2315 (2003).

59 Lahat N, Shapiro S, Karban A *et al.* Cytokine profile in coeliac disease. *Scand. J. Immunol.* 49, 441–446 (1999).

60 Forsberg G, Hernell O, Melgar S *et al.* Paradoxical coexpression of pro-inflammatory and downregulatory cytokines in intestinal T cells in childhood celiac disease. *Gastroenterology* 123, 667–678 (2002).

61 Nilsen EM, Jahnsen FL, Lundin KE *et al.* Gluten induces an intestinal cytokine response strongly dominated by interferon γ in patients with celiac disease. *Gastroenterology* 115, 551–563 (1998).

62 Woolley N, Mustalahti K, Maki M, Partanen J. Cytokine gene polymorphisms and genetic association with coeliac disease in the Finnish population. *Scand. J. Immunol.* 61, 51–56 (2005).

63 Beckett CG, Dell'Olio D, Kontakou M *et al.* Analysis of interleukin-4 and interleukin-10 and their association with the lymphocytic infiltrate in the small intestine of patients with coeliac disease. *Gut* 39, 818–823 (1996).

64 Cataldo F, Lio D, Marino V *et al.* Cytokine genotyping (TNF and IL-10) in patients with celiac disease and selective IgA deficiency. *Am. J. Gastroenterol.* 98, 850–856 (2003).

65 Hahn-Zoric M, Hytonen AM, Hanson LA, Nilsson LA, Padyukov L. Association of -1087 IL10 and -308 TNF A gene polymorphisms with serological markers of coeliac disease. *J. Clin. Immunol.* 23, 291–296 (2003).

66 Di Sabatino A, Ciccocioppo R, Carsetti R, Rosado MM, Cupelli F, Corazza GR. Defect of mucosal $CD4^+$ $CD25^+$ regulatory T-cells in coeliac disease. *11th International Symposium on Coeliac Disease, Belfast, UK* (2004) (Abstract).

67 Vaidya B, Pearce S. The emerging role of the CTLA-4 gene in autoimmune endocrinopathies. *Eur. J. Endocrinol.* 150, 619–626 (2004).

68 Ueda H, Howson JM, Esposito L *et al.* Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 423, 506–511 (2003).

69 Charles A Janeway PT, Mark Walport, Mark Shlomchik. In: *Immunobiology: The Immune System in Health and Disease.* Garland Publishing, NY, USA (2001).

70 Schuppan D. Current concepts of celiac disease pathogenesis. *Gastroenterology* 119, 234–242 (2000).

71 Vader W, Stepniak D, Kooy Y *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. USA* 100, 12390–12395 (2003).

72 Vader W, Kooy Y, Van Veelen P *et al.* The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* 122, 1729–1737 (2002).

73 Monteleone I, Monteleone G, Del Vecchio Blanco G *et al.* Regulation of the T helper cell Type 1 transcription factor T-bet in coeliac disease mucosa. *Gut* 53, 1090–5 (2004).

74 Salvati VM, MacDonald TT, Bajaj-Elliott M *et al.* Interleukin-18 and associated markers of T helper cell Type 1 activity in coeliac disease. *Gut* 50, 186–190 (2002).

75 Monteleone G, Pender SL, Alstead E *et al.* Role of interferon α in promoting T helper cell Type 1 responses in the small intestine in coeliac disease. *Gut* 48, 425–429 (2001).

76 Parrello T, Monteleone G, Cucchiara S *et al.* Upregulation of the IL-12 receptor β2 chain in Crohn's disease. *J. Immunol.* 165, 7234–7239 (2000).

77 Seegers D, Borm ME, van Belzen MJ *et al.* IL12B and IRF1 gene polymorphisms and susceptibility to celiac disease. *Eur. J. Immunogenet.* 30, 421–425 (2003).

78 Louka AS, Torinsson Naluai A, D'Alfonso S *et al.* The IL12B gene does not confer susceptibility to coeliac disease. *Tissue Antigens* 59, 70–72 (2002).

79 Przemioslo RT, Kontakou M, Nobili V, Ciclitira PJ. Raised pro-inflammatory cytokines interleukin-6 and tumour necrosis factor α in coeliac disease mucosa detected by immunohistochemistry. *Gut* 35, 1398–1403 (1994).

80 Westerholm-Ormio M, Garioch J, Ketola I, Savilahti E. Inflammatory cytokines in small intestinal mucosa of patients with potential coeliac disease. *Clin. Exp. Immunol.* 128, 94–101 (2002).

81 Kontakou M, Przemioslo RT, Sturgess RP, Limb AG, Ciclitira PJ. Expression of tumour necrosis factor-α, interleukin-6, and interleukin-2 mRNA in the jejunum of patients with coeliac disease. *Scand. J. Gastroenterol.* 30, 456–463 (1995).

82 Kontakou M, Przemioslo RT, Sturgess RP *et al.* Cytokine mRNA expression in the mucosa of treated coeliac patients after wheat peptide challenge. *Gut* 37, 52–57 (1995).

83 Polvi A, Maki M, Collin P, Partanen J. TNF microsatellite alleles a2 and b3 are not primarily associated with celiac disease in the Finnish population. *Tissue Antigens* 51, 553–555 (1998).

84 de la Concha EG, Fernandez-Arquero M, Vigil P *et al.* Celiac disease and TNF promoter polymorphisms. *Hum. Immunol.* 61, 513–517 (2000).

85 Garrote JA, Arranz E, Telleria JJ *et al.* TNF α and LT α gene polymorphisms as additional markers of celiac disease susceptibility in a DQ2-positive population. *Immunogenetics* 54, 551–555 (2002).

86 Louka AS, Lie BA, Talseth B *et al.* Coeliac disease patients carry conserved HLA-DR3-DQ2 haplotypes revealed by association of TNF alleles. *Immunogenetics* 55, 339–343 (2003).

87 Louka AS, Moodie SJ, Karell K *et al.* A collaborative European search for non-DQA1*05-DQB1*02 celiac disease loci on HLA-DR3 haplotypes: analysis of transmission from homozygous parents. *Hum. Immunol.* 64, 350–358 (2003).

88 Salvati VM, Mazzarella G, Gianfrani C *et al.* Recombinant human interleukin-10 suppresses gliadin dependent T cell activation in *ex vivo* cultured coeliac intestinal mucosa. *Gut* 54, 46–53 (2005).

89 Wapenaar MC, van Belzen MJ, Fransen JH *et al.* The interferon γ gene in celiac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. *J. Autoimmun.* 23, 183–190 (2004).

90 Breese EJ, Kumar P, Farthing MJ, MacDonald TT. Interleukin-2 and interferon-γ producing cells in the lamina propria in celiac disease. *Dig. Dis. Sci.* 39, 2243 (1994).

91 Veres G, Westerholm-Ormio M, Kokkonen J, Arato A, Savilahti E. Cytokines and adhesion molecules in duodenal mucosa of children with delayed-type food allergy. *J. Pediatr. Gastroenterol. Nutr.* 37, 27–34 (2003).

92 Kontakou M, Sturgess RP, Przemioslo RT *et al.* Detection of interferon γ mRNA in the mucosa of patients with coeliac disease by *in situ* hybridisation. *Gut* 35, 1037–1041 (1994).

93 al-Dawoud A, Nakshabendi I, Foulis A, Mowat AM. Immunohistochemical analysis of mucosal γ-interferon production in coeliac disease. *Gut* 33, 1482–1486 (1992).

94 Przemioslo RT, Lundin KE, Sollid LM, Nelufer J, Ciclitira PJ. Histological changes in small bowel mucosa induced by gliadin sensitive T lymphocytes can be blocked by anti-interferon γ antibody. *Gut* 36, 874–879 (1995).

95 Shuai K, Liu B. Regulation of JAK-STAT signalling in the immune system. *Nature Rev. Immunol.* 3, 900–911 (2003).

96 Mazzarella G, MacDonald TT, Salvati VM *et al.* Constitutive activation of the signal transducer and activator of transcription pathway in celiac disease lesions. *Am. J. Pathol.* 162, 1845–1855 (2003).

97 Salvati VM, MacDonald TT, del Vecchio Blanco G *et al.* Enhanced expression of interferon regulatory factor-1 in the mucosa of children with celiac disease. *Pediatr. Res.* 54, 312–318 (2003).

98 Jones SC, Trejdosiewicz LK, Banks RE *et al.* Expression of interleukin-6 by intestinal enterocytes. *J. Clin. Pathol.* 46, 1097–1100 (1993).

99 Ryan AW, Thornton JM, Brophy K *et al.* Chromosome 5q candidate genes in coeliac disease: genetic variation at IL4, IL5, IL9, IL13, IL17B and NR3C1. *Tissue Antigens* 65, 150–155 (2005).

100 Desreumaux P, Delaporte E, Colombel JF *et al.* Similar IL-5, IL-3, and

101 GM-CSF syntheses by eosinophils in the jejunal mucosa of patients with celiac disease and dermatitis herpetiformis. *Clin. Immunol. Immunopathol.* 88, 14–21 (1998).

102 Desreumaux P, Janin A, Colombel JF *et al.* Interleukin-5 messenger RNA expression by eosinophils in the intestinal mucosa of patients with coeliac disease. *J. Exp. Med.* 175, 293–296 (1992).

103 Sansonetti PJ. War and peace at mucosal surfaces. *Nature Rev. Immunol.* 4, 953–964 (2004).

104 Maiuri L, Auricchio S, Coletta S *et al.* Blockage of T-cell costimulation inhibits T-cell action in celiac disease. *Gastroenterology* 115, 564–572 (1998).

105 Maiuri L, Picarelli A, Boirivant M *et al.* Definition of the initial immunologic modifications upon *in vitro* gliadin challenge in the small intestine of celiac patients. *Gastroenterology* 110, 1368–1378 (1996).

106 Picarelli A, Maiuri L, Mazzilli MC *et al.* Gluten-sensitive disease with mild enteropathy. *Gastroenterology* 111, 608–616 (1996).

107 Mention JJ, Ben Ahmed M, Begue B *et al.* Interleukin-15: a key to disrupted intraepithelial lymphocyte homeostasis and lymphomagenesis in celiac disease. *Gastroenterology* 125, 730–745 (2003).

108 Martin-Pagola A, Ortiz L, Perez de Nanclares G *et al.* Analysis of the expression of MICA in small intestinal mucosa of patients with celiac disease. *J. Clin. Immunol.* 23, 498–503 (2003).

109 Fernandez L, Fernandez-Arquero M, Gual L *et al.* Triplet repeat polymorphism in the transmembrane region of the MICA gene in celiac disease. *Tissue Antigens* 59, 219–222 (2002).

110 Rueda B, Pascual M, Lopez-Nevot MA *et al.* Association of MICA-A5.1 allele with susceptibility to celiac disease in a family study. *Am. J. Gastroenterol.* 98, 359–362 (2003).

111 Lopez-Vazquez A, Rodrigo L, Fuentes D *et al.* MICA-A5.1 allele is associated with atypical forms of celiac disease in HLA-DQ2-negative patients. *Immunogenetics* 53, 989–991 (2002).

112 Lopez-Vazquez A, Rodrigo L, Fuentes D *et al.* MHC class I chain related gene A (MICA) modulates the development of coeliac disease in patients with the high risk heterodimer DQA1*0501/DQB1*0201. *Gut* 50, 336–340 (2002).

113 Diosdado B van Oort E, Wijmenga C. 'Celionomics': towards understanding the molecular pathology of celiac disease. *Clin. Chem. Lab. Med.* (2005) (In Press).

• **Summarizes the four molecular mechanisms that have been proposed as mediating villous atrophy in CD.**

114 Ciccocioppo R, Di Sabatino A, Parroni R *et al.* Increased enterocyte apoptosis and Fas–Fas ligand system in celiac disease. *Am. J. Clin. Pathol.* 115, 494–503 (2001).

115 Maiuri L, Ciacci C, Raia V *et al.* FAS engagement drives apoptosis of enterocytes of coeliac patients. *Gut* 48, 418–424 (2001).

116 Di Sabatino A, Ciccocioppo R, D'Alo S *et al.* Intraepithelial and lamina propria lymphocytes show distinct patterns of apoptosis whereas both populations are active in Fas based cytotoxicity in coeliac disease. *Gut* 49, 380–386 (2001).

117 Ehrmann J Jr., Kolek A, Kod'ousek R *et al.* Immunohistochemical study of the apoptotic mechanisms in the intestinal mucosa during children's coeliac disease. *Virchows Arch. A.* 442, 453–461 (2003).

118 Giovannini C, Matarrese P, Scazzocchio B *et al.* Wheat gliadin induces apoptosis of intestinal cells via an autocrine mechanism involving Fas-Fas ligand pathway. *FEBS Lett.* 540, 117–124 (2003).

119 Daum S, Bauer U, Foss HD *et al.* Increased expression of mRNA for matrix metalloproteinases-1 and -3 and tissue inhibitor of metalloproteinases-1 in intestinal biopsy specimens from patients with coeliac disease. *Gut* 44, 17–25 (1999).

120 Daum S, Bauer U, Foss HD *et al.* Expression of matrix metalloprotease-1 and collagen I mRNA in biopsies from patients with celiac disease. *Ann. NY Acad. Sci.* 859, 254–257 (1998).

121 Salmela MT, MacDonald TT, Black D *et al.* Upregulation of matrix metalloproteinases in a model of T cell mediated tissue injury in the gut: analysis by gene array and *in situ* hybridisation. *Gut* 51, 540–547 (2002).

122 Pender SL, Tickle SP, Docherty AJ *et al.* A major role for matrix metalloproteinases in T cell injury in the gut *J. Immunol.* 158, 1582–1590 (1997).

123 Rivabene R, Mancini E, De Vincenzi M. *In vitro* cytotoxic effect of wheat gliadin-derived peptides on the Caco-2 intestinal cell line is associated with intracellular oxidative imbalance: implications for coeliac disease. *Biochim. Biophys. Acta.* 1453, 152–160 (1999).

124 Elli L, Dolfini E, Bardella MT. Gliadin cytotoxicity and *in vitro* cell cultures. *Toxicol. Lett.* 146, 1–8 (2003).

125 Stahlberg MR, Hietanen E, Maki M. Mucosal biotransformation rates in the small intestine of children. *Gut* 29, 1058–1063 (1988).

126 Odetti P, Valentini S, Aragno I *et al.* Oxidative stress in subjects affected by celiac disease. *Free Radic. Res.* 29, 17–24 (1998).

127 Lavy A, Ben Amotz A, Aviram M. Increased susceptibility to undergo lipid peroxidation of chylomicrons and low-density lipoprotein in celiac disease. *Ann. Nutr. Metabol.* 37, 68–74 (1993).

128 Maiuri MC, De Stefano D, Mele G *et al.* Gliadin increases iNOS gene expression in interferon-γ-stimulated RAW 264.7 cells through a mechanism involving NF-κB. *Naunyn Schmiedebergs Arch. Pharmacol.* 368, 63–71 (2003).

129 Bach SP, Renehan AG, Potten CS. Stem cells: the intestinal stem cell as a paradigm. *Carcinogenesis* 21, 469–476 (2000).

130 Salvati VM, Bajaj-Elliott M, Poulsom R *et al.* Keratinocyte growth factor and coeliac disease. *Gut* 49, 176–181 (2001).

131 Senger S, Luongo D, Maurano F *et al.* Intranasal administration of a recombinant α-gliadin downregulates the immune response to wheat gliadin in DQ8 transgenic mice. *Immunol. Lett.* 88, 127–134 (2003).

132 Choi K, Siegel M, Piper JL *et al.* Chemistry and biology of dihydroisoxazole derivatives: selective inhibitors of human transglutaminase 2. *Chem. Biol.* 12, 469–475 (2005).

133 De Laurenzi V, Melino G. Gene disruption of tissue transglutaminase. *Mol. Cell Biol.* 21, 148–155 (2001).

134 Nanda N, Iismaa SE, Owens WA *et al.* Targeted inactivation of Gh/tissue transglutaminase II. *J. Biol. Chem.* 276, 20673–20678 (2001).

135 Choy EH, Scott DL, Kingsley GH *et al.* Control of rheumatoid arthritis by oral tolerance. *Arthritis Rheum.* 44, 1993–1997 (2001).

136 Thompson DJ, Barron KS, Whitcup SM, Robinson MR. The safety and efficacy of chicken Type II collagen on uveitis associated with juvenile rheumatoid arthritis. *Ocul. Immunol. Inflamm.* 10, 83–91 (2002).

Affiliations

- *Begoña Diosdado, MD*
  University Medical Centre, Complex Genetics Section, Stratenum 2.117, Department of Biomedical Genetics, PO Box 85060, 3508 AB Utrecht, The Netherlands
  Tel.: +31 30 253 8790
  Fax: +31 30 253 8479
  b.diosdado@med.uu.nl
- *Cisca Wijmenga, PhD*
  University Medical Centre, Complex Genetics Section, Stratenum 2.117, Department of Biomedical Genetics, PO Box 85060, 3508 AB Utrecht, The Netherlands
  Tel.: +31 30 253 8427
  Fax: +31 30 253 8479
  t.n.wijmenga@med.uu.nl

Chapter 2

"Coelionomics": towards understanding the molecular pathology of coeliac disease

Clin Chem Lab Med. 2005;43(7):685-95

"Coelionomics": towards understanding the molecular pathology of coeliac disease[1)]

Begoña Diosdado*, Erica van Oort* and Cisca Wijmenga**

Complex Genetics Section, DBG-Department of Medical Genetics, University Medical Centre, Utrecht, The Netherlands

Abstract

Coeliac disease (CD) is an inflammatory disorder of the small intestine characterised by a permanent intolerance to gluten-derived peptides. When gluten-derived peptides reach the lamina propria in CD patients, they provoke specific changes in the mucosa of their small intestine. Although the susceptibility to CD is strongly determined by environmental gluten, it is clearly a common genetic disorder. Important genetic factors for CD are the *HLA-DQ* genes located in the MHC region on chromosome 6 [*HLA-DQ2* (95%) or *HLA-DQ8* (~5%) heterodimers]. So far, the only treatment for CD consists of a life-long gluten-free diet. A key question in CD is why the gluten-derived peptides are resistant to further breakdown by endogenous proteases and how, in turn, they can activate a harmful immune response in the lamina propria of genetically predisposed individuals. Four mechanisms, namely apoptosis, oxidative stress, matrix metalloproteinases and dysregulation of proliferation and differentiation, are thought to play a role in the pathophysiology of CD. Whether the genes involved in these four mechanisms play a causative role in the development of the villous atrophy or are, in fact, a consequence of the disease process is unknown. In this review we summarise these mechanisms and discuss their validity in the context of current insights derived from genetic, genomic and molecular studies. We also discuss future directions for research and the therapeutic implications for patients.

Keywords: apoptosis; coeliac disease; matrix metalloproteinases; oxidative stress; proliferation and differentiation.

Introduction

Coeliac disease (CD) (OMIM 212750) is an inflammatory disorder of the small intestine characterised by a permanent intolerance to gluten-derived peptides. Gluten is a dietary protein present in wheat, rye and barley. When gluten-derived peptides reach the lamina propria in CD patients, they provoke specific changes in the mucosa of their small intestine (Figure 1). These histological changes were described by Marsh in 1992 and he classified the disease into three stages according to the mucosa morphology: Marsh I stage, characterised by lymphocytic infiltration in the mucosa; Marsh II stage, in which crypt hyperplasia develops in addition to the Marsh I features; and Marsh III stage, with villous atrophy together with the Marsh II features (1).

CD patients can be clinically either symptomatic or asymptomatic. Typical CD symptoms include chronic diarrhoea, malabsorption, and a failure to thrive in children. Atypical symptoms are often seen, such as anaemia or skin lesions. CD is a common autoimmune disease with a prevalence of 1:100–200 in European and American populations (2), but it is rare in Japanese, Chinese and native Africans. The disease is more common among women, with a 3:1 ratio, and is associated with Down (3) and Turner syndromes (4), as well as other autoimmune disorders such as type I diabetes mellitus, psoriasis and thyroid diseases (5, 6).

When untreated, CD patients have a higher risk of developing T-cell lymphoma, infertility and osteoporosis. So far, the only treatment for CD consists of a life-long gluten-free diet. This diet reverses the clinical and histological symptoms and probably prevents other complications from developing. Whether the treatment protects CD patients from developing other autoimmune disorders is not entirely clear yet.

Although the susceptibility to CD is strongly determined by environmental gluten, it is clearly a common genetic disorder. The concordance rate amongst monozygotic twins is estimated to be 86%, compared to 20% in dizygotic twins (7), implying a large degree of heritability. Although the segregation of the disease does not follow a Mendelian pattern of inheritance, there is strong familial clustering: approximately 10% of first-degree relatives of CD patients are affected, conferring a 20-fold increased risk over the general population (~0.5%). Important genetic factors for CD are the *HLA-DQ* genes located in the MHC region on chromosome 6, as CD is mainly limited to individuals who can form either *HLA-DQ2* (95%) or *HLA-DQ8* (~5%) heterodimers. Interestingly,

---

[1)]Presented at the Second Santorini Prospective Conference "From Human Genetic Variations to Prediction of Risks and Responses to Drugs and to the Environment", Santorini, Greece, Sep 30–Oct 4, 2004.

* Both authors contributed equally to this paper.

**Corresponding author: Prof. Cisca Wijmenga, Complex Genetics Section, Department of Biomedical Genetics, Stratenum 2.117, University Medical Centre Utrecht, P.O. Box 85060 AT, 3508 AB Utrecht, The Netherlands
Phone: +31-30 253 8427, Fax: +31-30 253 8479,
E-mail: t.n.wijmenga@med.uu.nl

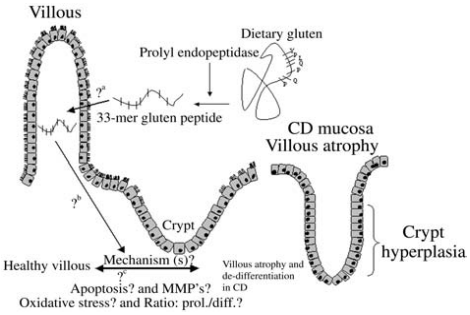


Figure 1 Basic scheme representing the healthy villous (left) and villous atrophy (right) seen under the influence of gluten peptides in coeliac disease (CD) patients. How the 33-mer gluten peptide passes through the enterocytes is still unknown (a). Subsequently, these peptides lead to activation of the immune response (native and/or adaptive) in CD patients (b). These actions eventually direct a combination of the proposed pathomechanisms (c) resulting in villous atrophy and de-differentiation in CD patients ingesting gluten. All three subsequent steps still need to be elucidated.

however, approximately 30% of the general population also express HLA-DQ2 or HLA-DQ8 molecules, implying that non-HLA genes must also be important determinants of disease development.

Disease pathology

A key question in CD is why the gluten-derived peptides are resistant to further breakdown by endogenous proteases and how, in turn, they can activate a harmful immune response in the lamina propria of genetically predisposed individuals. It has been known for some time that the gluten peptides are presented by HLA-DQ2 and HLA-DQ8 molecules to naive T-cells, and that when these are activated they elicit an inflammatory response by releasing pro-inflammatory cytokines. These cytokines provoke damage to the epithelial barrier and release of the enzyme tissue transglutaminase. This enzyme deaminates the glutamine residues of the gluten peptides into negatively charged glutamic acid residues, thereby enhancing their binding to HLA-DQ2 or HLA-DQ8 molecules and amplifying the Th1-mediated immune response (8). More recently, it has been proposed that gluten itself might be directly responsible for the immune response, either by activation of the innate immune response mediated by interleukin-15 (IL-15), or by a direct effect of gluten on the enterocytes (9–13).

It has been shown that CD4+ T-cells play a central role in controlling the immune response to gluten that causes the immunopathology in CD, although the actual mechanisms responsible for tissue damage have only been partially characterised. Different hypotheses have been put forward to explain the villous atrophy seen in CD, including a cytotoxic effect of the intraepithelial lymphocytes leading to apoptosis of enterocytes (12–26), tissue damage by alterations in the oxidative metabolism (27–35), destruction of the connective tissue by imbalanced expression of matrix metalloproteinases (MMPs) (36–43), and a block of terminally differentiation of the cells of the intestinal crypts (44–49) (Figure 1). In this review we aim to summarise these mechanisms and discuss their validity in the context of current insights derived from genetic, genomic and molecular studies. We also discuss future directions for research and the therapeutic implications for patients.

Tissue remodelling: potential mechanisms leading to structural changes in CD

The role of cytotoxicity leading to apoptosis of villous enterocytes

Induced cell death or apoptosis of the enterocytes has been proposed as one of the major mechanisms responsible for villous atrophy in CD. Several studies have investigated which pathways mediate apoptosis in the enterocytes of CD patients and two mechanisms have been proposed. The first is that apoptosis in CD is mediated by intraepithelial lymphocytes (IELs) and/or lamina propria lymphocytes (LPLs) (12–17); the second suggests that toxic domains in gliadin peptides are capable of directly stimulating apoptosis of enterocytes (18, 19).

FAS ligand (FASL), a cell surface molecule belonging to the tumour necrosis factor family, binds to its receptor FAS, thereby inducing apoptosis of FAS-bearing cells. Stress or damage induces the expression of FAS, which converts these cells into targets for the killer lymphocytes that express FASL. FAS and FASL are involved in T-cell-mediated cytotoxicity.

The involvement of apoptosis by IELs and LPLs has been studied by immunohistochemical detection of FAS and FASL antigens, as well as by DNA fragmentation studies using TdT-mediated dUTP nick-end labelling (TUNEL) assays. All these studies (12–17), except the one by Ehrmann et al. (17), reported increased numbers of $FASL^+$ IELs and FAS^+ enterocytes in untreated CD patients compared to treated CD patients and controls. These results advocate exacerbation of the FAS system in CD patients, which in turn can cause tissue destruction. Moreover, an increased number of $TUNEL^+$ enterocytes have been observed in CD patients, and these were positively correlated with the FAS^+ cells (14–17). To gather support for these observations, a large number of complementary experiments were performed (summarised in supplementary Table S1 on website: http://humgen.med.uu.nl/publications/CD/Diosdado2005_1/index.html).

Recent findings have added further evidence that apoptosis in CD can be mediated by IELs and increased IELs is a hallmark of active CD. Works by Meresse et al. (12) and Hue et al. (13) add additional insight into the molecular mechanisms of the cytotoxicity by IELs that has been shown to occur in CD patients. Meresse et al. (12) showed that IELs, under conditions of deregulated IL-15 expression, transform into so-called NK-like LAK (''lymphokine-activated kill-

Table 1 Expression and genetic studies performed on genes implicated in pathogenesis of CD.

Process involved	Gene	Chromosome location	Mapped to linkage region	Population	Association studies	Reported in mRNA expression studies
Apoptosis	*FAS*	1p36	No		Severity of villus atrophy (92)	No
Apoptosis	*FASL*	1q23	No		ND	No
Apoptosis	*Granzyme B*	14q11.2	No		ND	No
Apoptosis	*Perforin*	10q22	No		ND	No
Apoptosis	*B-cl2*	18q21.33	No		ND	No
Apoptosis	*NKG2D*	12p13.2	No		ND	No
Apoptosis	*MICA*	6p21.3	CELIAC1 (56)	European/ American	HLADQ2 (93, 94)	No change (95)
Apoptosis	*DAP12*	19q13.1	No		ND	No
Apoptosis	*IL-15*	4q31	No		ND	Increased (96, 97)
MMPs-related	*MMP1*	11q22.3	No		Negative (44)	No
MMPs-related	*MMP3*	11q22.3	No		Negative (44)	No
MMPs-related	*MMP9*	20q11.2-q13.1	No		ND	No
MMPs-related	*MMP10*	11q21-23	No		ND	No
MMPs-related	*MMP12*	11q21-23	No		ND	No
MMPs-related	*TIMP1*	Xp11.3-p11.23	No		ND	No
Oxidative stress	*Apolipoproteins*	(12q13.3)	No			
Oxidative stress	*NF-κB*	4q24	No		ND	No
Oxidative stress	*STAT-1a*	2q32.2	CELIAC3 (57, 58)	Finnish/ Swedish/ Norwegian	No	No
Oxidative stress	*IRF-1*	5q31.1	CELIAC2 (56, 61, 62)	Italian/ European consortium	Negative (98)	No
Oxidative stress	*IFNG*	12q14	No		Negative (99)	Increased (99–100)
Oxidative stress	*iNOS*	17q11.2	No		ND	No
Oxidative stress	*Glutathione-S-transferase*	6p21.1			ND	No
Oxidative stress/ MPPs-related	*TNF-α*	6p21.3			Associated with microsatellite allele TNFb3 (102)	Increased (100)
Differentiation/ proliferation	*TM4SF4*	3q25	No		ND	Decreased (49)
Peptide cleavage	*PREP*	6q22	6q21-22 (67)	Dutch	Negative (76)	Increased in presence of gluten (76)
Peptide deamidation	*tTG*	20q12	No		Negative (103)	Increased (104)

ers'') cells that express the NKG2D receptor. Interestingly, the ligand of NKG2D, MICA, is up-regulated in active CD patients. Hence, activation of the NKG2D-MICA cascade mediates a cytolytic effect on the epithelial cells through activation of DAP12, PIE3 and ERK pathways (12, 13). These findings suggest that IL-15 is a key mediator in this process and that the interaction between MICA-NKG2D is critical for killing the enterocytes.

The direct cytotoxic effect of gliadin on enterocytes was noted in 1976 by Weiser and Douglas (20). This hypothesis was experimentally tested in several epithelial cell lines using two- and three-dimensional cell cultures. The results demonstrated that gliadin has a direct cytotoxic effect on monolayers of cells (21, 22), resulting in cytoskeleton rearrangements (23–25) and a decrease in cell viability, as well as inhibiting cell growth (19, 26).

Over the years a large set of functional data has implied a role for apoptosis in villous atrophy, whether this mechanism is solely responsible for the observed damage is still being debated, partly because the different studies are difficult to compare. Some were conducted on cell lines, whereas others used biopsies, and the specificity of the frequently used TUNEL assay is disputed (27). Genetic studies have so far not been able to detect linkage or associations for the genes implicated in apoptosis, or have not yet been performed (e.g., FAS, FASL, granzyme B). An exception is MICA on 6p21.3, which has been located in the CELIAC1 region (Table 1).

Involvement of oxidative stress and nitric oxide in villous atrophy

Increased free-radical formation is referred to as oxidative stress and, together with a decreased reductive potential, it has the capacity to induce cell damage or even cell death. In CD it has been reported that gliadin can induce oxidative stress responses (28), for example, increasing concentrations of peptic-tryptic-digested bread wheat gliadin have been shown to alter the oxidative balance in colonic carcinoma cell cultures (CaCo-2) (28). In these experiments the digested

gliadin was incubated with undifferentiated CaCo-2 cells, after which the following parameters were measured: 4-hydroxy-2(E)-nonenal (4-HNE), reduced and oxidised glutathione (GSH) content, protein sulphhydryl groups (SH) and plasma membrane damage by annexin V. These parameters showed an increasing 4-HNE content, reductions in GSH content, the disappearance of protein thiols and a positive reaction to annexin V, suggesting an oxidative imbalance upon gliadin incubation (28).

Another study using Lovo cell cultures instead of CaCo-2 also showed a gliadin concentration-dependent reduction in GSH content of up to 34% with respect to controls (29). In a recent study, the cytotoxic effect of pepsin-pancreatic-digested bread wheat gliadin was tested on three-dimensional cell cultures. It was observed that 7 days after stimulation the cells showed reduced viability and morphology indicative of cell damage (19). Apart from cell lines, diminished epoxide hydrolase and glutathione peroxidase activity could be demonstrated in biopsies taken from CD patients with mild to severe villous atrophy (MII-MIII) (30). Other stress markers, such as carbonyl groups on proteins and lipid peroxidation, showed significantly higher levels when derived from CD patients compared to controls, whereas antioxidants such as α-tocopherol (vitamin E) and lipoproteins (cholesterol and apolipoproteins) were reduced (31, 32). Nitric oxide synthase (NOS) exists as three separate isoforms, neuronal NOS (nNOS), endothelial NOS (eNOS) and inducible NOS (iNOS), which have also been implicated in mucosal damage. The expression of iNOS in the small intestine was found to correlate with high levels of nitric oxide (NO) products in the urine of children with active CD (33).

Recently, it was demonstrated that gluten peptides could enhance the mRNA expression of iNOS, although the mechanism is not yet understood, and thereby increasing NO production by mouse macrophages stimulated with interferon (IFN)γ. Three transcription factors, NF-κB, STAT-1a and IRF-1, also seem to be involved in the enhanced expression of iNOS (34).

This suggests that gluten, by an as yet unknown mechanism, plays a role in the activation of NF-κB and the transcription of other pro-inflammatory genes (34). Subsequent mucosal damage is then generated by these inflammatory molecules, together with an increase in NO, which can act as a free radical (35). Another study obtained biopsies from patients with irritable bowel syndrome (IBS) and CD patients to analyse the NOS activity in both the lamina propria (eNOS) and isolated enterocytes (iNOS). Newly diagnosed CD patients showed highly elevated iNOS activity compared to patients on a gluten-free diet. Although iNOS activity was increased in untreated patients, the authors could not associate this finding with elevated cGMP levels. The most relevant action of NO is the activation of soluble guanylate cyclase, with a subsequent increase in cGMP levels. Nor could a correlation be found between detected nitrated proteins and iNOS activity in the enterocyte population (36). Nevertheless, these data suggest that, in an environment with elevated IFNγ, gluten can somehow induce up-regulation of TNF-α and iNOS expression, both of which can play a role in the mechanism leading to villous atrophy.

Involvement of matrix metalloproteinases (MMPs) in villous atrophy

Proteinases such as MMPs were thought to be mainly responsible for turnover and degradation of connective tissue proteins. However, more recently, it has been suggested that this is not their main role, as they were found to act on non-matrix proteins such as chemokines and anti-microbial peptides, thereby facilitating the activity of these proteins. Whether the exact role of MMPs is to promote or attenuate the state of inflammation has not yet been elucidated (37).

In CD the extracellular matrix degradation of MMPs might be implicated in the disease mechanism leading to villous atrophy. It has been shown that mRNA expression levels of MMP-1 and MMP-3 and a tissue inhibitor of metalloproteinases (TIMP-1) were increased in biopsies of patients with untreated CD. After treatment with a gluten-free diet, the expression returned to normal, while the TIMP-1 levels remained elevated. Neither of these expression levels influenced the expression of collagen 1, which remained unaltered in all test subjects (38). Another study used T-cells in explants of human foetal gut and gene array technology to investigate the changes induced upon T-cell activation with pokeweed mitogen. The T-cell-mediated tissue injury resulted in up-regulation of MMP-1, -3, -9, -10 and -12, whereas the TIMPs (TIMP 1 to 4) were unchanged or down-regulated. This finding suggests that a change in the balance between proteases and anti-proteases leads to mucosal degradation (39). Furthermore, the elevated release of TNF-α upon T-cell activation, also associated with mucosal injury, has been shown to elevate MMPs expression, leading to an environment promoting mucosal degradation (40, 41).

Elevation of expression of MMP-1 and -3 has also been implicated as playing a role in dermatitis herpetiformis (DH), a gluten-related skin disorder. Although gastrointestinal symptoms in DH patients are rare, over 80% show gluten-sensitive enteropathy manifested by crypt hyperplasia and partial villous atrophy, suggesting a common disease aetiology between DH and CD (42).

Association studies with functional polymorphisms in the *MMP-1* and *MMP-3* gene promoter regions located on chromosome 11 (43) were performed in a collection of CD patients from 327 Norwegian and Swedish families and 160 Norwegian controls. Although the 11q22.2 region has repeatedly shown evidence for linkage in CD, no significant evidence to support association of the two polymorphisms of *MMP-1* and *MMP-3* with CD could be found (44) (Table 1).

MMPs are normally produced at very low concentrations, and have a primary role in repairing tissue injury and tissue remodelling. It might therefore be

expected that, in the studies described above, elevated levels of MMPs would be found, as it is inevitable that tissue injury is induced when using human gut explants and biopsies. However, any type of tissue remodelling (villous atrophy, recovery of the mucosa) will have an influence on the levels of MMPs. Therefore, it seems reasonable to assume that MMP-1 and -3 are involved in the pathogenesis of CD, but whether elevated activity in the gut of CD patients is a result of gene variation with expression modifications has not been confirmed. Thus far, results regarding the expression levels of MMPs in CD patients are probably rather a consequence of CD pathogenesis than a causative event.

Distortion in the proliferation/differentiation ratio of villous cells

To date, little is known about the mucosal changes seen in CD and the mechanisms involved in the reappearance of normal villous-crypt architecture after treating CD patients with a gluten-free diet. Another unexplained characteristic is the hyperplasia of the crypt cells. Wright et al. (45) developed a method to measure the number of proliferating cells per crypt and the crypt/cell production rate in control (villous) mucosa and flat (avillous) mucosa (CD patients), as well as in patients suffering from DH. Correcting for the size of the proliferating populations in control and flat mucosa, they obtained values for cell cycle times of 45 and 22 h, respectively. They stated that the mitotic duration is the same, but that the time between subsequent divisions is shortened by half in CD. In flat mucosa of adult CD patients, this resulted in a three to four-fold increase in total crypt cell populations compared to controls. If calculated as cell production rates per hour, the adult control group showed a rate of 25 cells per hour, and in adult CD and DH this increased to 150 cells per hour (46, 47). Based on these studies, the hyperplasia of untreated CD patients seems to be accounted for by an increase in the cell division rate (48). So far, there is speculation that the hyperplasia observed is a response to a loss of villous architecture in the pathological cell; however, a plausible explanation for the loss of differentiated cells in the villous and consequent loss of villous architecture has not been given. Wright et al. (47) have also shown that the maturation compartment in both adult CD and DH showed an absolute and relative increase in size compared to control values. The total number of maturing cells per crypt increased from 220 in controls to 1370 in adult CD patients. However, either the increased population of proliferating cells is still not adequate to replace the cell loss or the maturing cells do not completely develop into terminally differentiated enterocytes.

Diosdado et al. (49) performed microarray experiments on biopsy sections from healthy controls and CD patients with total villous atrophy (Marsch III lesions). By comparing Marsh III biopsies with control biopsies, 109 genes were identified that differed significantly in expression levels. A large number of these genes have functions in proliferation and differentiation pathways, and might be important for the correct development of crypt-villous units. The analyses also revealed 120 differentially expressed genes when patients on a gluten-free diet were compared with those exposed to gluten. These results confirm the overall observation that, in the presence of gluten, the small intestine of CD patients converts to a state of increased cell proliferation (48, 49). Thus, overall the microarray data suggest an enhanced state of proliferation and a diminished differentiation status in the small intestine of CD patients. One of the genes that may be involved in the switch from proliferation to differentiation is the TransMembrane 4 Super Family 4 gene (*TM4SF4*) (50). In untreated CD patients, TM4SF4 was found to be two-fold down-regulated in microarrays (49), but an up to nine-fold down-regulation in expression was observed in RT-PCR. Protein expression levels of TM4SF4 analysed by immunohistochemistry also showed a gradual decrease in expression (van Oort et al., in preparation). Although a higher state of proliferation in CD patients is observed (crypt hyperplasia), this does not seem to compensate for the lack of terminally differentiated enterocytes in the villi, ultimately leading to villous atrophy. A malfunction in the switch from proliferation to differentiation could explain the abundance of proliferating cells and the lack of fully differentiated cells maintaining the villous structure. However, an increase in the velocity in which cell migration occurs could also result in less terminally differentiated cells. It is therefore possible that inhibition of the process of differentiation and/or the degree of cell movement is directed by a polymorphism in an as yet unknown ''CD gene'', or is secondary to activation of the immune response. In this second scenario, two growth factors regulated by the immune response have been studied in CD: keratinocyte growth factor, which increases proliferation and has been shown to be expressed at high levels in CD patients (51), and transforming growth factor-β, which induces differentiation and patchy staining in the villi of CD patients, while normal controls show strong staining (52, 53).

Towards identifying the molecular pathways contributing to CD

The mechanisms described above probably all contribute, in some way or other, to the tissue damage observed in CD patients after dietary ingestion of gluten. It is likely that these pathways interact with each other, as it is known that the *MICA* gene is transcriptionally regulated by H_2O_2-induced oxidative stress (54), and that MMPs can affect cell-cell adhesion and consequently cell proliferation and differentiation (55). However, we still do not know what the exact molecular pathways are that trigger and induce further progression of the process of inflammation and subsequent tissue remodelling. The identification of genes causally related to the disease will be instrumental in indicating key components in these molec-

ular pathways. In addition, gene expression profiling and proteomics may complement genetic research strategies, since genomic approaches can point to the consequences and reveal molecular insight into the disease-associated pathways. The integration of both genetic and genomic information will be pivotal for our complete understanding of the molecular basis of the disease pathology. This knowledge can then be used to improve diagnosis and prevention of CD, and help develop novel therapeutic strategies.

Identification of causative genes in CD

Two approaches can be followed to identify non-HLA genes in CD: (i) genetic association studies of candidate genes, and (ii) genome-wide screening. Genetic association studies have mainly been performed with genes from immunological and inflammatory response pathways. The most consistent result was obtained with the *CTLA4* gene locus, although there is still controversy about whether *CTLA4* itself or a neighbouring gene is responsible for this positive association (56). Genome-wide screening approaches do not require insight into the disease mechanism and are therefore not hypothesis-driven. To date, 11 genome-wide screens have been performed in Scandinavian (57–60), Italian (61, 62), Irish (63), UK (64), North American (60, 65), and Dutch (66, 67) populations. In addition, a meta-analysis has been conducted by the EU-CD consortium (56). All these studies showed significant linkage to the HLA region on chromosome 6p (CELIAC1), but only a few studies showed significant linkage outside this region. In the Finnish population, significant linkage was observed to chromosomes 2 (CELIAC3 locus) (60) and 15 (68), while in the Dutch population a locus on the short arm of chromosome 19 (CELIAC4 locus) has been implicated (67). A large number of suggestive linkages have been reported that still await replication in independent populations.

The key question is how to go from a linked region to a disease-causing gene. The most obvious strategy is to perform genetic association studies with single nucleotide polymorphisms (SNPs) in such a way that all the genetic information is captured. Recent insight into the organisation of genetic variation along the genome suggested that a significant proportion of the human genome is contained within so-called ''haplotype blocks''. These blocks are genomic regions showing strong genetic association (i.e., linkage disequilibrium, LD) between pairwise SNPs (69). The goal of the International HapMap Project is to characterise these patterns of LD (70), while understanding LD is expected to aid the discovery of genes that influence complex human diseases, since local disequilibrium patterns may help to identify common polymorphisms involved in complex disease. Interestingly, the strong LD between SNPs also means that only a small subset of the approximately 11 million common SNPs present in the human genome needs to be tested to indirectly detect hidden, disease-associated variants based on LD (71). A small-scale association study was actually utilised for the 3.5 Mb CELIAC4 locus in the Dutch population and revealed association to SNPs within a single gene, *MYO9b* (67). Five associated SNPs do indeed combine into a single haplotype that is significantly increased on comparing cases (38.8%) and controls (30.9%) ($p < 0.02$) (unpublished results). To prove that this gene does indeed contribute to CD susceptibility, the observed association needs to be replicated in independent cohorts, the extent of the association's boundaries need to be determined, the associated region needs to be sequenced to identify the true disease-causing variant, and functional studies need to be performed. Recent examples from the literature show that it is a cumbersome process to identify genetic variants involved in complex diseases, because the variants themselves are not exclusive to the diseased population, as is also evident from the *HLA-DQ2* genes in CD.

Identifying the molecular pathways in CD: gene expression profiling

Since the completion of the human genome project in 2003, genomics and proteomics technologies have developed exponentially. CD is a particularly suitable disease to study using these new technologies, since the site of the lesions where the pathogenesis takes place is relatively easy to reach and the main environmental factor, gluten, can be manipulated in vivo, as well as in vitro.

Identifying genes with altered expression due to their involvement in disease aetiology and/or pathology is facilitated by microarray hybridisations or quantitative real-time polymerase chain reaction (RT-PCR). So far, only three gene expression profiling studies using gene arrays have been performed in CD. One of these was conducted on human foetal explants and focused on the differential expression of MMPs, showing up-regulation of MMP1, -3, -9, -10 and -12 (39). The other two studies (49, 75) were performed on intestinal biopsies from CD patients and controls. Their aim was to define which genes were differentially expressed between villous atrophy and normal mucosa, and before and after treatment on a gluten-free diet. Although both studies had a similar experimental set-up and were both performed with biopsy material, the results did not show much overlap. This lack of correlation might be explained by differences in the platforms that were used (hybridisation-based membranes vs. cDNA-spotted microarrays) and the subsequent data analyses. Since microarray experiments generate millions of data points, much attention needs to be paid to the statistical analysis [reviewed by Quackenbush (72)]. In addition, it is important to deposit the minimum information about a microarray experiment (MIAME) (73) into ArrayExpress – a public repository for microarray gene expression data (74). In the study performed by Juuti-Uusitalo et al. (75), up-regulation of genes related to transport pathways was observed, whereas in the study performed by Diosdado et al. (49), mainly genes involved in proliferation-differentiation pathways, such as *TM4SF4* and *EphB3*, were identified

after comparing CD to control biopsies. Interestingly, the study by Diosdado et al. (49) also identified differential expression of prolyl-endopeptidase (PREP), an enzyme involved in gluten metabolism, when the authors compared treated to untreated CD patients. This *PREP* gene is interesting not only from a functional point of view – as it might be important in proteolytic cleavage of the proline-rich gluten peptides – but also from a genetic point of view. The *PREP* gene is located in the long arm of chromosome 6, under a linkage peak that showed suggestive linkage to CD in the Dutch population (67). However, extensive association studies excluded *PREP* as a primary gene in CD pathogenesis (76).

The overall picture from combining the two studies points to changes in similar processes: activated Th1 response, enhanced cell proliferation, reduced epithelial differentiation, recruitment of γδ T-cells, B-cells and macrophages, and a lack of changes in MMPs. No indication of the involvement of apoptosis genes was found. Although these results merely reflect the histological changes observed in CD lesions, we expect a more detailed understanding of the molecular processes underlying these changes to reveal the true mechanism behind these changes.

Identifying the molecular pathways in CD: proteomics

In the past decade, microarray technology has become widely available for studying transcriptional regulation. Genome-wide screens for protein functions have, until recently, been carried out using cDNA libraries and immobilisation of the expressed proteins on a membrane. However, these techniques are not suitable for high-throughput screening of thousands of proteins. With new high-throughput molecular techniques, an effort has been made to develop microarray-based methods for the study of protein function, but facilitating these studies on a genome-wide level is still complicated. Furthermore, these assays should be able to contain full-length and correctly folded proteins. The techniques facilitating these requirements have been developed for a subset of proteins and peptides, allowing the study of protein-protein interactions and the interaction between protein kinases and their substrates (77). Instead of protein arrays, antibody arrays are also being developed (78). Other high-throughput studies involve mass spectrometry (79), allowing the screening of hundreds of samples with detection limits in the femtomolar range. The search for biomarkers related to specific diseases such as CD in body fluids such as plasma and serum, as well as in ground tissues, is now possible (80, 81). A few years ago, (functional-) proteomics related to CD was used to identify acyl-acceptor and acyl-donor substrates for tissue transglutaminase (tTG) in CaCo-2 cell lines. This led to the identification of new tTG protein substrates (82) and the local site of its enzymatic activity (83). Studies such as these can lead to better understanding of the still unclear function of tTG, as well as its exact role in CD. Another study identified three new CD autoantigens, ATP synthase β-chain and two charge variants of enolase α using sera and biopsy specimens of active CD patients compared to controls (84). However, the usefulness of the proteins identified for CD screening, diagnosis or follow-up has yet to be determined. These approaches could be useful in identifying biomarkers in plasma of CD patients, and in simplifying screening and diagnostics. Despite the potential usefulness of proteomics, many problems still have to be resolved. For example, the abundance of human serum albumin and immunoglobulin G in plasma and serum obscures the detection of hundreds of other proteins present in smaller quantities. Another problem is that many proteins give rise to different splice variants or show post-translational modifications such as glycosylation. For many proteins, these modifications are still unknown. Furthermore, the complexity of proteomics, 20,000 genes (microarrays) vs. an unknown number of proteins, is still problematic.

Summary and future prospects

Over the past few years, our knowledge on the pathophysiology of CD has increased substantially. We now know that an individual must express *HLA-DQ2/8* to become a CD patient, and we know how gluten modification by tTG can enhance the T-cell response. However, the mechanism behind gluten resistance to proteolytic breakdown, the passage through the epithelial layer, and the connection between the inflammatory response and subsequent tissue remodelling are still poorly understood. Although it is evident that key cytokines such as IFNγ and TNF-α play an important role in this process, it is not known how they are associated with crypt hyperplasia and villous atrophy.

Four mechanisms, namely apoptosis, oxidative stress, MMPs and dysregulation of proliferation and differentiation, are thought to play a role in the pathophysiology of CD. Whether the genes involved in these four mechanisms play a causative role in the development of the villous atrophy or are, in fact, a consequence of the disease process is unknown. Recent advances in gene technology will help to answer these questions. These include microarrays, which compare gene expression levels, and proteomics, which looks at protein expression levels and isoforms. So far, the results obtained from microarray studies do not show changes in expression of genes involved in apoptotic pathways, oxidative stress, or in *MMP* genes. Consistently, negative results were obtained in the genetic studies performed for some of these candidate genes (Table 1). Furthermore, the genetic regions identified by genome-wide screens do not contain obvious functional candidate genes involved in any of these three mechanisms (Table 1). These findings may imply that these genes are a consequence of the disease rather than a causal factor. However, microarray studies performed on CD biopsies do confirm the involvement of proliferation and differentiation pathways in CD. Hence, studying

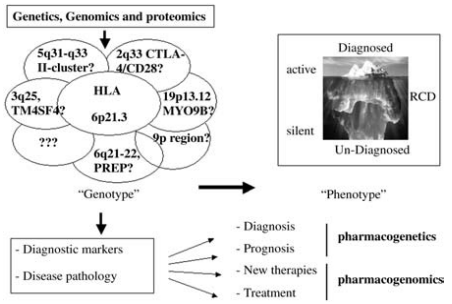


Figure 2 Outline of the different disciplines involved in unravelling coeliac disease (CD) pathology. Genetic and genomic (microarray) studies have implicated a number of putative genes and loci in CD. In future, increasing insight will lead to the application of these genes as diagnostic markers. Moreover, insight into the disease pathways will increase knowledge of the disease pathology. These insights will eventually help to resolve the CD iceberg, as only one in seven CD patients is currently diagnosed. Development of diagnostic and prognostic markers can also help to determine patient-specific therapies, as patients seem to have different genetic backgrounds. Currently, therapy is limited to a life-long gluten-free diet, which is not sufficient for all CD patients; a few do not respond to this diet and become refractory (RCD patients). The development of new techniques and integration of different fields (proteomics, pharmacogenomics, bioinformatics) may eventually lead to unravelling of the disease pathogenesis, which is expected to lead to new therapies specific for the primary genetic defects.

changes in these novel pathways might be fruitful in uncovering which molecular mechanisms are responsible for the pathology of CD, and could consequently direct the search for candidate genes in the genetically linked and associated regions. It is expected that an approach that integrates genetics, genomics and proteomics will contribute to the discovery of causative genes for CD (Figure 2).

Another area that will further aid in the search for candidate genes is the field of bioinformatics. For example, molecular networks developed by the integration of gene expression, gene annotation, and protein-protein interaction data may contribute to identifying molecular pathways correlated with diseases (85–87).

Finally, identifying genes related to CD susceptibility will be important in developing prognostic scoring rules that can be used to identify patients at risk of developing CD with high specificity and sensitivity. Currently, the only gold standard for diagnosing CD patients is by taking small intestinal biopsies, which involves an invasive procedure. Since CD is largely under-diagnosed, with only one diagnosed patient to seven undiagnosed patients, there is an urgent need for molecular diagnostics. Such genes may offer the possibility of studying phenotype/genotype correlations, as well as the response to a gluten-free diet. In addition, genome information from both genomics and genetic studies is expected to play an important role in drug development. Insight into the molecular pathways may also open up new avenues for therapy (Figure 2), as this information will enhance our success in identifying potential targets for drugs (pharmacogenomics) (88). Although a gluten-free diet is relatively easy to follow, it might not offer protection against all the potential complications of CD. Recently, some alternatives to the gluten-free treatment have been proposed. These include enzymotherapy to digest gluten into peptides that are too small to be presented by HLA class II molecules (89), vaccinotherapy to immunosuppress the adaptive immune response (90) and genetic modification of gluten-rich grains to reduce their toxicity for CD patients (91). When evaluating new therapeutic interventions, it might be sensible to consider the genetic background of the individuals, as CD patients with different genetic make-ups may respond differently (Figure 2). Genetic information might be instrumental in guiding clinical trials by helping to predict which patients are likely to be potential responders or hyper-responders. Alternatively, the genotype associated with non-responsiveness or toxicity (pharmacogenetics) can be mapped using genome-wide SNP association studies (88). Hence, predictive SNP profiles will allow testing of patients before starting therapeutic intervention. Similarly, SNP profiles might become available to predict the progression of disease, allowing interventions to be tailored to the individual patient.

Acknowledgements

We thank Jackie Senior and Martin Wapenaar for critically reading the manuscript, and the Netherlands Organisation for Medical and Health Research (NWO) grant 902-22-094 and ZON-MW program grant 912-02-028 for financial support. CDC, the Celiac Disease Consortium, is an innovative cluster approved by the Netherlands Genomics Initiative and partially funded by a Dutch government grant (BSIK03009).

References

1. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). Gastroenterology 1992;102:330–54.
2. Book L, Zone JJ, Neuhausen SL. Prevalence of celiac disease among relatives of sib pairs with celiac disease in US families. Am J Gastroenterol 2003;98:377–81.
3. Carnicer J, Farre C, Varea V, Vilar P, Moreno J, Artigas J. Prevalence of coeliac disease in Down's syndrome. Eur J Gastroenterol Hepatol 2001;13:263–7.
4. Bonamico M, Pasquino AM, Mariani P, Danesi HM, Culasso F, Mazzanti L, et al. Prevalence and clinical picture of celiac disease in Turner syndrome. J Clin Endocrinol Metab 2002;87:5495–8.
5. Sategna Guidetti C, Solerio E, Scaglione N, Aimo G, Mengozzi G. Duration of gluten exposure in adult coeliac disease does not correlate with the risk for autoimmune disorders. Gut 2001;49:502–5.
6. Collin P, Kaukinen K, Valimaki M, Salmi J. Endocrinological disorders and celiac disease. Endocr Rev 2002; 23:464–83.

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;50:624–8.
8. Schuppan D. Current concepts of celiac disease pathogenesis. Gastroenterology 2000;119:234–42.
9. Maiuri L, Picarelli A, Boirivant M, Coletta S, Mazzilli MC, De Vincenzi M, et al. Definition of the initial immunologic modifications upon in vitro gliadin challenge in the small intestine of celiac patients. Gastroenterology 1996;110: 1368–78.
10. Maiuri L, Ciacci C, Ricciardelli I, Vacca L, Raia V, Auricchio S, et al. Association between innate response to gliadin and activation of pathogenic T cells in coeliac disease. Lancet 2003;362:30–7.
11. Godkin A, Jewell D. The pathogenesis of celiac disease. Gastroenterology 1998;115:206–10.
12. 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.
13. 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.
14. Ciccocioppo R, Di Sabatino A, Parroni R, D'Alo S, Pistoia MA, Doglioni C, et al. Cytolytic mechanisms of intraepithelial lymphocytes in coeliac disease (CoD). Clin Exp Immunol 2000;120:235–40.
15. Maiuri L, Ciacci C, Raia V, Vacca L, Ricciardelli I, Raimondi F, et al. FAS engagement drives apoptosis of enterocytes of coeliac patients. Gut 2001;48:418–24.
16. Di Sabatino A, Ciccocioppo R, D'Alo S, Parroni R, Millimaggi D, Cifone MG, et al. Intraepithelial and lamina propria lymphocytes show distinct patterns of apoptosis whereas both populations are active in Fas based cytotoxicity in coeliac disease. Gut 2001;49:380–6.
17. Ehrmann J Jr, Kolek A, Kod'ousek R, Zapletalova J, Lisova S, Murray PG, et al. Immunohistochemical study of the apoptotic mechanisms in the intestinal mucosa during children's coeliac disease. Virchows Arch 2003; 442:453–61.
18. Giovannini C, Matarrese P, Scazzocchio B, Vari R, D'Archivio M, Straface E, et al. Wheat gliadin induces apoptosis of intestinal cells via an autocrine mechanism involving Fas-Fas ligand pathway. FEBS Lett 2003; 540:117–24.
19. Dolfini E, Elli L, Dasdia T, Bufardeci B, Colleoni MP, Costa B, et al. In vitro cytotoxic effect of bread wheat gliadin on the LoVo human adenocarcinoma cell line. Toxicol In Vitro 2002;16:331–7.
20. Weiser MM, Douglas AP. An alternative mechanism for gluten toxicity in coeliac disease. Lancet 1976;1:567–9.
21. Rocca E, Paganuzzi Stammati A, Zampaglioni F, Zucco F. Effects of gliadin-derived peptides from bread and durum wheats on in vitro cultures of human cell lines. Implications for coeliac disease pathogenesis. Toxicol Lett 1983;16:331–8.
22. Hudson DA, Cornell HJ, Purdham DR, Rolles CJ. Non-specific cytotoxicity of wheat gliadin components towards cultured human cells. Lancet 1976;1:339–41.
23. Clemente MG, De Virgiliis S, Kang JS, Macatagney R, Musu MP, Di Pierro MR, et al. Early effects of gliadin on enterocyte intracellular signalling involved in intestinal barrier function. Gut 2003;52:218–23.
24. Sjolander A, Magnusson KE. Effects of wheat germ agglutinin on the cellular content of filamentous actin in intestine 407 cells. Eur J Cell Biol 1988;47:32–5.
25. Glenney JR Jr, Glenney P. Comparison of Ca^{++}-regulated events in the intestinal brush border. J Cell Biol 1985;100:754–63.
26. Dolfini E, Elli L, Ferrero S, Braidotti P, Roncoroni L, Dasdia T, et al. Bread wheat gliadin cytotoxicity: a new three-dimensional cell model. Scand J Clin Lab Invest 2003;63:135–41.
27. Walker-Smith J. 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.
28. Rivabene R, Mancini E, De Vincenzi M. In vitro cytotoxic effect of wheat gliadin-derived peptides on the Caco-2 intestinal cell line is associated with intracellular oxidative imbalance: implications for coeliac disease. Biochim Biophys Acta 1999;1453:152–60.
29. Elli L, Dolfini E, Bardella MT. Gliadin cytotoxicity and in vitro cell cultures. Toxicol Lett 2003;146:1–8.
30. Stahlberg MR, Hietanen E, Maki M. Mucosal biotransformation rates in the small intestine of children. Gut 1988;29:1058–63.
31. Odetti P, Valentini S, Aragno I, Garibaldi S, Pronzato MA, Rolandi E, et al. Oxidative stress in subjects affected by celiac disease. Free Radic Res 1998;29:17–24.
32. Lavy A, Ben Amotz A, Aviram M. Increased susceptibility to undergo lipid peroxidation of chylomicrons and low-density lipoprotein in celiac disease. Ann Nutr Metab 1993;37:68–74.
33. van Straaten EA, Koster-Kamphuis L, Bovee-Oudenhoven IM, van der Meer R, Forget PP. Increased urinary nitric oxide oxidation products in children with active coeliac disease. Acta Paediatr 1999;88:528–31.
34. Maiuri MC, De Stefano D, Mele G, Iovine B, Bevilacqua MA, Greco L, et al. Gliadin increases iNOS gene expression in interferon-gamma-stimulated RAW 264.7 cells through a mechanism involving NF-kappa B. Naunyn Schmiedebergs Arch Pharmacol 2003;368:63–71.
35. Liu RH, Hotchkiss JH. Potential genotoxicity of chronically elevated nitric oxide: a review. Mutat Res 1995; 339:73–89.
36. Murray IA, Daniels I, Coupland K, Smith JA, Long RG. Increased activity and expression of iNOS in human duodenal enterocytes from patients with celiac disease. Am J Physiol Gastrointest Liver Physiol 2002;283: G319–26.
37. Parks WC, Wilson CL, Lopez-Boado YS. Matrix metalloproteinases as modulators of inflammation and innate immunity. Nat Rev Immunol 2004;4:617–29.
38. Daum S, Bauer U, Foss HD, Schuppan D, Stein H, Riecken EO, et al. Increased expression of mRNA for matrix metalloproteinases-1 and -3 and tissue inhibitor of metalloproteinases-1 in intestinal biopsy specimens from patients with coeliac disease. Gut 1999;44:17–25.
39. Salmela MT, MacDonald TT, Black D, Irvine B, Zhuma T, Saarialho-Kere U, et al. Upregulation of matrix metalloproteinases in a model of T cell mediated tissue injury in the gut: analysis by gene array and in situ hybridisation. Gut 2002;51:540–7.
40. Pender SL, Fell JM, Chamow SM, Ashkenazi A, MacDonald TT. A p55 TNF receptor immunoadhesin prevents T cell-mediated intestinal injury by inhibiting matrix metalloproteinase production. J Immunol 1998;160:4098–103.
41. Pender SL, Tickle SP, Docherty AJ, Howie D, Wathen NC, MacDonald TT. A major role for matrix metalloproteinases in T cell injury in the gut. J Immunol 1997; 158:1582–90.
42. Salmela MT, Pender SL, Reunala T, MacDonald T, Saarialho-Kere U. Parallel expression of macrophage metalloelastase (MMP-12) in duodenal and skin lesions of patients with dermatitis herpetiformis. Gut 2001; 48:496–502.
43. Ye S. Polymorphism in matrix metalloproteinase gene promoters: implication in regulation of gene expres-

sion and susceptibility of various diseases. Matrix Biol 2000;19:623–9.
44. Louka AS, Stensby EK, Ek J, Gudjonsdottir AH, Ascher H, Sollid LM. Coeliac disease candidate genes: no association with functional polymorphisms in matrix metalloproteinase 1 and 3 gene promoters. Scand J Gastroenterol 2002;37:931–5.
45. Wright NA, Morley AR, Appleton DR, Marks JM, Douglas AP, Watson AJ. Measurement of cell production rate in the human small bowel. Pathol Microbiol (Basel) 1973;39:251–3.
46. Wright N, Watson A, Morley A, Appleton D, Marks J, Douglas A. The cell cycle time in the flat (avillous) mucosa of the human small intestine. Gut 1973; 14:603–6.
47. Wright N, Watson A, Morley A, Appleton D, Marks J. Cell kinetics in flat (avillous) mucosa of the human small intestine. Gut 1973;14:701–10.
48. Savidge TC, Walker-Smith JA, Phillips AD. Intestinal proliferation in coeliac disease: looking into the crypt. Gut 1995;36:321–3.
49. 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.
50. Wice BM, Gordon JI. A tetraspan membrane glycoprotein produced in the human intestinal epithelium and liver that can regulate cell density-dependent proliferation. J Biol Chem 1995;270:21907–18.
51. Salvati VM, Bajaj-Elliott M, Poulsom R, Mazzarella G, Lundin KE, Nilsen EM, et al. Keratinocyte growth factor and coeliac disease. Gut 2001;49:176–81.
52. Bach SP, Renehan AG, Potten CS. Stem cells: the intestinal stem cell as a paradigm. Carcinogenesis 2000;21:469–76.
53. Lionetti P, Pazzaglia A, Moriondo M, Azzari C, Resti M, Amorosi A, et al. Differing patterns of transforming growth factor-beta expression in normal intestinal mucosa and in active celiac disease. J Pediatr Gastroenterol Nutr 1999;29:308–13.
54. Yamamoto K, Fujiyama Y, Andoh A, Bamba T, Okabe H. Oxidative stress increases MICA and MICB gene expression in the human colon carcinoma cell line (CaCo-2). Biochim Biophys Acta 2001;1526:10–2.
55. Hojilla CV, Mohammed FF, Khokha R. Matrix metalloproteinases and their tissue inhibitors direct cell fate during cancer development. Br J Cancer 2003; 89:1817–21.
56. 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;11:828–34.
57. Holopainen P, Mustalahti K, Uimari P, Collin P, Maki M, Partanen J. Candidate gene regions and genetic heterogeneity in gluten sensitivity. Gut 2001;48:696–701.
58. Naluai AT, Nilsson S, Gudjonsdottir AH, Louka AS, Ascher H, Ek J, et al. Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for celiac disease on chromosomes 5 and 11. Eur J Hum Genet 2001;9:938–44.
59. Liu J, Juo SH, Holopainen P, Terwilliger J, Tong X, Grunn A, et al. Genomewide linkage analysis of celiac disease in Finnish families. Am J Hum Genet 2002; 70:51–9.
60. Rioux JD, Karinen H, Kocher K, McMahon SG, Karkkainen P, Janatuinen E, et al. Genomewide search and association studies in a Finnish celiac disease population: identification of a novel locus and replication of the HLA and CTLA4 loci. Am J Med Genet 2004; 130A:345–50.
61. Greco L, Corazza G, Babron MC, Clot F, Fulchignoni-Lataud MC, Percopo S, et al. Genome search in celiac disease. Am J Hum Genet 1998;62:669–75.
62. Greco L, Babron MC, Corazza GR, Percopo S, Sica R, Clot F, et al. Existence of a genetic risk factor on chromosome 5q in Italian coeliac disease families. Ann Hum Genet 2001;65:35–41.
63. Zhong F, McCombs CC, Olson JM, Elston RC, Stevens FM, McCarthy CF, et al. An autosomal screen for genes that predispose to celiac disease in the western counties of Ireland. Nat Genet 1996;14:329–33.
64. King AL, Yiannakou JY, Brett PM, Curtis D, Morris MA, Dearlove AM, et al. A genome-wide family-based linkage study of coeliac disease. Ann Hum Genet 2000;64:479–90.
65. Neuhausen SL, Feolo M, Camp NJ, Farnham J, Book L, Zone JJ. Genome-wide linkage analysis for celiac disease in North American families. Am J Med Genet 2002;111:1–9.
66. van Belzen MJ, Vrolijk MM, Meijer JW, Crusius JB, Pearson PL, Sandkuijl LA, et al. A genomewide screen in a four-generation Dutch family with celiac disease: evidence for linkage to chromosomes 6 and 9. Am J Gastroenterol 2004;99:466–71.
67. 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.
68. Woolley N, Holopainen P, Ollikainen V, Mustalahti K, Maki M, Kere J, et al. A new locus for coeliac disease mapped to chromosome 15 in a population isolate. Hum Genet 2002;111:40–5.
69. Cardon LR, Abecasis GR. Using haplotype blocks to map human complex trait loci. Trends Genet 2003; 19:135–40.
70. The International HapMap Project. Nature 2003; 426:789–96.
71. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, et al. Whole-genome patterns of common DNA variation in three human populations. Science 2005;307:1072–9.
72. Quackenbush J. Computational analysis of microarray data. Nat Rev Genet 2001;2:418–27.
73. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. Nat Genet 2001;29:365–71.
74. Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, Coulson R, et al. ArrayExpress – a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 2005;33:D553–5.
75. 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.
76. 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 celiac disease population. Am J Physiol Gastrointest Liver Physiol 2005 (doi:10.1152/ajpgi.00056.2005). In press.
77. MacBeath G, Schreiber SL. Printing proteins as microarrays for high-throughput function determination. Science 2000;289:1760–3.
78. Kwon Y, Han Z, Karatan E, Mrksich M, Kay BK. Antibody arrays prepared by cutinase-mediated immobilization on self-assembled monolayers. Anal Chem 2004;76:5713–20.
79. Bogdanov B, Smith RD. Proteomics by FTICR mass spectrometry: top down and bottom up. Mass Spectrom Rev 2005;24:168–200.
80. Journet A, Ferro M. The potentials of MS-based subproteomic approaches in medical science: the case of lyso-

somes and breast cancer. Mass Spectrom Rev 2004;23:393–442.
81. Yeo S, Roh GS, Kim DH, Lee JM, Seo SW, Cho JW, et al. Quantitative profiling of plasma peptides in asthmatic mice using liquid chromatography and mass spectrometry. Proteomics 2004;4:3308–17.
82. Orru S, Caputo I, D'Amato A, Ruoppolo M, Esposito C. Proteomics identification of acyl-acceptor and acyl-donor substrates for transglutaminase in a human intestinal epithelial cell line. Implications for celiac disease. J Biol Chem 2003;278:31766–73.
83. Ruoppolo M, Orru S, D'Amato A, Francese S, Rovero P, Marino G, et al. Analysis of transglutaminase protein substrates by functional proteomics. Protein Sci 2003;12:1290–7.
84. Stulik J, Hernychova L, Porkertova S, Pozler O, Tuckova L, Sanchez D, et al. Identification of new celiac disease autoantigens using proteomic analysis. Proteomics 2003;3:951–6.
85. Vidal M. A biological atlas of functional maps. Cell 2001;104:333–9.
86. Ishii N, Robert M, Nakayama Y, Kanai A, Tomita M. Toward large-scale modeling of the microbial cell for computer simulation. J Biotechnol 2004;113:281–94.
87. Franke L, van Bakel H, Diosdado B, van Belzen M, Wapenaar M, Wijmenga C. TEAM: a tool for the integration of expression, and linkage and association maps. Eur J Hum Genet 2004;12:633–8.
88. Roses AD. Pharmacogenetics and drug development: the path to safer and more effective drugs. Nat Rev Genet 2004;5:645–56.
89. 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.
90. Senger S, Luongo D, Maurano F, Mazzeo MF, Siciliano RA, Gianfrani C, et al. Intranasal administration of a recombinant alpha-gliadin down-regulates the immune response to wheat gliadin in DQ8 transgenic mice. Immunol Lett 2003;88:127–34.
91. Vader LW, Stepniak DT, Bunnik EM, Kooy YM, de Haan W, Drijfhout JW, et al. Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. Gastroenterology 2003;125: 1105–13.
92. Wu J, Alizadeh BZ, Veen TV, Meijer JW, Mulder CJ, Pena AS. Association of FAS (TNFRSF6)-670 gene polymorphism with villous atrophy in coeliac disease. World J Gastroenterol 2004;10:717–20.
93. Lopez-Vazquez A, Rodrigo L, Fuentes D, Riestra S, Bousono C, Garcia-Fernandez S, et al. MHC class I chain related gene A (MICA) modulates the development of coeliac disease in patients with the high risk heterodimer DQA1*0501/DQB1*0201. Gut 2002;50:336–40.
94. Rueda B, Pascual M, Lopez-Nevot MA, Koeleman BP, Ortega E, Maldonado J, et al. Association of MICA-A5.1 allele with susceptibility to celiac disease in a family study. Am J Gastroenterol 2003;98:359–62.
95. Martin-Pagola A, Ortiz L, Perez de Nanclares G, Vitoria JC, Castano L, Bilbao JR. Analysis of the expression of MICA in small intestinal mucosa of patients with celiac disease. J Clin Immunol 2003;23:498–503.
96. Maiuri L, Ciacci C, Auricchio S, Brown V, Quaratino S, Londei M. Interleukin 15 mediates epithelial changes in celiac disease. Gastroenterology 2000;119:996–1006.
97. Mention JJ, Ben Ahmed M, Begue B, Barbe U, Verkarre V, Asnafi V, et al. Interleukin 15: a key to disrupted intraepithelial lymphocyte homeostasis and lymphomagenesis in celiac disease. Gastroenterology 2003; 125:730–45.
98. Seegers D, Borm ME, van Belzen MJ, Mulder CJ, Bailing J, Crusius JB, et al. IL12B and IRF1 gene polymorphisms and susceptibility to celiac disease. Eur J Immunogenet 2003;30:421–5.
99. Wapenaar MC, van Belzen MJ, Fransen JH, Sarasqueta AF, Houwen RH, Meijer JW, et al. The interferon gamma gene in celiac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. J Autoimmun 2004;23:183–90.
100. Lahat N, Shapiro S, Karban A, Gerstein R, Kinarty A, Lerner A. Cytokine profile in coeliac disease. Scand J Immunol 1999;49:441–6.
101. Monteleone I, Monteleone G, Del Vecchio Blanco G, Vavassori P, Cucchiara S, MacDonald TT, et al. Regulation of the T helper cell type 1 transcription factor T-bet in coeliac disease mucosa. Gut 2004;53:1090–5.
102. McManus R, Moloney M, Borton M, Finch A, Chuan YT, Lawlor E, et al. Association of celiac disease with microsatellite polymorphisms close to the tumor necrosis factor genes. Hum Immunol 1996;45:24–31.
103. 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.
104. Esposito C, Paparo F, Caputo I, Porta R, Salvati VM, Mazzarella G, et al. Expression and enzymatic activity of small intestinal tissue transglutaminase in celiac disease. Am J Gastroenterol 2003;98:1813–20.

Received January 13, 2005, accepted May 19, 2005

Part II

GENETIC PROFILING STUDIES IN COELIAC DISEASE

Chapter 3

A microarray screen for novel candidate genes in coeliac disease pathogenesis

Gut. 2004 July;53(7):944-51

A microarray screen for novel candidate genes in coeliac disease pathogenesis

B Diosdado, M C Wapenaar, L Franke, K J Duran, M J Goerres, M Hadithi, J B A Crusius, J W R Meijer, D J Duggan, C J J Mulder, F C P Holstege, C Wijmenga

Gut 2004;**53**:944–951. doi: 10.1136/gut.2003.018374

See end of article for authors' affiliations

Correspondence to: Professor C Wijmenga, Department of Biomedical Genetics, Stratenum 2. 117, University Medical Centre Utrecht, PO Box 80030, 3508 TA Utrecht, the Netherlands; t.n.wijmenga@med.uu.nl

Background and aims: The causative molecular pathways underlying the pathogenesis of coeliac disease are poorly understood. To unravel novel aspects of disease pathogenesis, we used microarrays to determine changes in gene expression of duodenal biopsies.

Methods: cDNA microarrays representing 19 200 genes were used to compare gene expression profiles of duodenal biopsies from 15 coeliac disease patients with villous atrophy (Marsh III) and seven control individuals with normal biopsies (Marsh 0). In addition, the specific effect of gluten was studied by comparing the expression profiles of Marsh III lesions of seven patients exposed to gluten with four patients on a gluten free diet.

Results: Comparing Marsh III with Marsh 0 lesions identified 109 genes that differed significantly ($p<0.001$) in expression levels between patients and controls. A large number of these genes have functions in proliferation and differentiation pathways and might be important for correct development of crypt-villous units. Alterations in these pathways may lead to the characteristic hyperplasia and villous atrophy seen in coeliac disease. The analyses also revealed 120 differentially expressed genes ($p<0.005$) when comparing patients on a gluten free diet with those exposed to gluten. These genes further strengthen our observation of increased cell proliferation in the presence of gluten.

Conclusions: Our study provides new candidate genes in the pathogenesis of coeliac disease. Based on our results, we hypothesise that villous atrophy in coeliac disease patients is due to failure in cell differentiation. These genes are involved in pathways not previously implicated in coeliac disease pathogenesis and they may provide new targets for therapy.

Coeliac disease (OMIM 212750) is a chronic inflammatory enteropathy caused by lifelong intolerance to gluten in genetically predisposed individuals. When gluten, the main protein present in wheat, rye, and barley,[1] is ingested by individuals expressing human leucocyte antigen (HLA)-DQ2 and/or -DQ8 heterodimers,[2] it provokes a misdirected immune response in the small intestine where gluten is absorbed. This immune response leads to a series of histological changes resulting in lymphocytosis, crypt hyperplasia, and villous atrophy. These histological abnormalities, which were documented by Marsh in 1992,[3] are characteristic of this disease and explain some of the clinical symptoms of coeliac disease patients. Treatment of coeliac disease consists of a lifelong gluten free diet, resulting in clinical recovery and histological normalisation of the intestine.

Coeliac disease is a multifactorial disorder for which both genetic and environmental factors are required for the disease phenotype to develop.[4] The molecular pathways responsible for the disease pathogenesis are currently only partly defined.[5] So far only one genetic factor has been identified, the HLA-DQA and -DQB gene cluster on chromosome 6.[2] Although many attempts are underway to identify the other genetic factors involved in coeliac disease, traditional genetic mapping studies face serious limitations in identifying the full repertoire of susceptibility genes due to the small contribution of each individual susceptibility gene and the complex interplay between genetic and environmental factors. Genomics technologies, such as cDNA microarrays, are expected to provide additional insight into the molecular and cellular aspects of diseases.

Coeliac disease is a unique disease as the environmental factor, gluten, that triggers the disease is known and can be easily manipulated in vivo as well as in vitro. Moreover, the site of the lesion is easily accessible and small intestinal biopsies need to be taken from coeliac disease patients as part of the standard diagnostic procedure. In order to expand our understanding of the pathogenesis of coeliac disease, we set out to perform cDNA microarray analysis on a set of well characterised duodenum biopsies from coeliac disease patients with classic histopathology (Marsh III (MIII)) and control individuals (Marsh 0 (M0)). Microarray analysis is a powerful technique that allows the study of the level and pattern of expression of thousands of genes simultaneously.[6]

We first thought that comparing MIII with M0 biopsies might lead to the discovery of genes involved in the immune response to gluten and the long term tissue destruction seen in the small intestine of coeliac disease patients. Subsequently, we investigated the molecular changes that occur in MIII biopsies by comparing a collection of MIII biopsies from individuals who were on a gluten free diet with those on a gluten containing diet (that is, exposed to gluten).

We report two sets of genes that have not previously been associated with the pathogenesis of coeliac disease. Our results imply a role for novel candidate genes involved in the maintenance of the intestinal villi. As some of these genes map to chromosomal regions implicated in genetic mapping studies, they may represent possible causal candidate genes.

The combination of well characterised intestinal biopsies and cDNA microarrays is a unique method of studying the molecular and cellular events taking place in the initiation

Abbreviations: HLA, human leucocyte antigen; M, Marsh; Th, T helper; RT-PCR, reverse transcription-polymerase chain reaction; IFN, interferon; IL, interleukin; APC, antigen presenting cells

and further progression towards the mucosal transformation seen in coeliac disease. It has not yet been applied to this disease.

MATERIALS AND METHODS

Patients

Intestinal biopsies from 15 coeliac disease patients were included in the study. All biopsies showed MIII histology and were evaluated by one pathologist (JWRM) according to the modified UEGW criteria.[7]

Seven patients (patients 1–7, table 1) suspected of having coeliac disease were on a gluten containing diet when MIII villous atrophy was histologically confirmed.

Another four coeliac disease patients (patients 8–11, table 1) were on a gluten free diet for a year and reported total clinical recovery although their intestines still showed MIII characteristics. A slight histological improvement was reported. In addition, biopsies were taken from the last group of four patients (patients 12–15, table 1), who were refractory coeliac disease type I patients.[8] Biopsies of these four refractory coeliac disease type I patients showed MIII histology and no clinical improvement[8] despite the fact that they were on a strict gluten free diet. A team of dieticians monitored compliance to the diet in all patients on a gluten free diet.

In addition, duodenum biopsies from seven individuals who had an endoscopic examination for other reasons were used as control samples. The histology of these tissue biopsies was completely normal (M0). Characteristics of the coeliac disease patients and controls included in this study are summarised in table 1.

The Medical Ethical Committee of the University Medical Centre of Utrecht approved the study. All patients, or the parents of paediatric patients, included in the study gave written informed consent.

Biopsy sampling and RNA isolation

For each individual, two to three biopsies (15–20 mg) were taken from the proximal duodenum by spike forceps endoscopy. Fresh tissue samples were snap frozen and stored in liquid nitrogen. Frozen biopsies were homogenised in TRIzol with glass beads of 1 mm diameter using a Mini-BeadBeater (BioSpec Products, Inc, Bartlesville, Oklahoma, USA), and total RNA was isolated using TRIzol (Gibco/Life Technologies, Rockville, Maryland, USA) following the manufacturer's protocol. The quality and quantity of the RNA samples was determined using a 2100 Agilent Bioanalyzer (Agilent Technologies, Palo Alto, California, USA). Biopsies yielded an average of 3.6 µg of RNA/mg of tissue.

Microarray hybridisation

The seven controls and 15 coeliac disease samples were analysed by hybridisation screening of cDNA microarrays obtained from the University Health Network of Toronto, Ontario (Canada). Two different releases of slide sets, 19k2 and 19k3, containing 19 200 genes printed in duplicate on two glass slides were used for the experiments.

First strand cDNA was prepared from 10 µg of total RNA from biopsies and labelled with Cy3, as described by Van de Peppel and colleagues.[9] A surgical specimen from the small bowel was used to generate a Cy5 labelled cDNA probe for use as a reference in all coeliac disease and control hybridisations. Between 200 and 300 ng of labelled cDNA from the biopsies and reference tissues were used for a single set of slides for overnight hybridisation. Slides were scanned in a ScanArray 4000 XL (Packard BioScience, Boston, Massachusetts, USA).

An MIAME compliant[10] downloadable dataset and full details of the protocols are available at http://humgen.med.uu.nl/publications/CD/Microarrays/Diosdado_1.html or http://www.ebi.ac.uk/arrayexpress/, experiment accession No E-MEXP-42.

Image and data analysis

The scanned images were subjected to image analysis using the Dearray suite[11] for IPLab SPECTRUM software (Scanalytics, Inc., Fairfax, Virginia, USA). The software, in a

Table 1 Characteristics of coeliac disease, refractory gluten free diet type I patients, and controls included in the study

Sample	Age/Sex	Diagnosis	Biopsy status‡	GFD¶	HLA typing	Slide type§
1	18/F	CD patient*	MIIIb	No	DQ2+/DQ8−	19k2
2	3/M	CD patient	MIIIb	No	DQ2+/DQ8−	19k3
3	54/F	CD patient	MIIIc	No	DQ2+/DQ8−	19k3
4	33/F	CD patient	MIIIa	No	DQ2+/DQ8−	19k3
5	2/F	CD patient	MIIIc	No	DQ2+/DQ8−	19k3
6	79/F	CD patient	MIIIb	No	DQ2+/DQ8−	19k3
7	38/M	CD patient	MIIIa	No	DQ2+/DQ8−	19k2
8	61/M	CD patient	MIIIb	Yes	DQ2+/DQ8−	19k3
9	40/M	CD patient	MIIIb	Yes	DQ2+/DQ8−	19k3
10	61/F	CD patient	MIIIa	Yes	DQ2+/DQ8−	19k3
11	70/M	CD patient	MIIIa	Yes	DQ2+/DQ8−	19k3
12	51/F	RCD type I patient†	MIIIa	Yes	DQ2+/DQ8−	19k3
13	29/F	RCD type I patient	MIIIb-c	Yes	DQ2+/DQ8+	19k3
14	84/F	RCD type I patient	MIIIa	Yes	DQ2+/DQ8−	19k3
15	72/F	RCD type I patient	MIIIa	Yes	DQ2-/DQ8−	19k3
16	32/F	Control	M0	No	DQ2+/DQ8+	19k2
17	40/F	Control	M0	No	DQ2-/DQ8−	19k3
18	51/F	Control	M0	No	n/d	19k3
19	60/F	Control	M0	No	n/d	19k3
20	36/F	Control	M0	No	DQ2-/DQ8−	19k2
21	32/F	Control	M0	No	n/d	19k2
22	33/F	Control	M0	No	DQ2-/DQ8+	19k2

*CD patient, coeliac disease patient; †RCD patient, refractory gluten free diet type I patient.
‡Histology was reviewed by the same pathologist (JWRM). Biopsies were classified according to the UEGW criteria.
¶GFD, gluten free diet.
§Two releases of microarray slides, 19k2 and 19k3, from the University Health Network, Toronto, Ontario (Canada) were used for hybridisations.

semi-automatic manner, identifies the fluorescent spots, subtracts the local background, and determines a quality score for each spot based on the spot's intensity, size, local background, and uniformity of intensity within the spot.[12] Based on these parameters, a quality score is assigned to each individual spot. All spots with a quality score >0.01 were selected for further processing.

To correct for differences in the efficiencies between the Cy3 and Cy5 channel, a method of global normalisation was applied. The quantified signal intensity for the entire array in both the Cy3 and Cy5 channels was averaged and equalised by applying a normalisation factor. Subsequently, this normalisation correction was applied to each individual spot and the red to green ratio was calculated.[12] As genes are spotted in duplicate, the average of the [10]log of the signal of each of the two copies of the same gene was calculated only when the quality score for both genes was >0.01. When only one copy of the gene had a quality score >0.01, only the level of expression of that gene copy was used.

Data analysis was performed with the GeneSpring package, version 4.2.1 (Silicon Genetics, Redwood City, California, USA). Genes whose expression was significantly different between two groups of biopsy samples were selected from the genes present on the slides by applying a Welch *t* test. Cluster analysis was performed using a supervised cluster algorithm.

Gene analysis

To further define the biological function of the selected genes, a homemade Java tool database was developed (Franke *et al*, in press) for data storage, gene classification, and gene analysis. This database also contains information on selected genes such as GeneBank accession No, Locus Link ID, chromosomal location, Emsembl ID, Unigene information, Gene Ontology, and GeneCards ID.

Data validation by real time RT-PCR

A selection of genes that showed altered expression in the microarray analysis was re-examined by real time reverse transcription-polymerase chain reaction (RT-PCR) to validate the changes observed in an independent manner. Firstly, cDNA was generated from 1 µg of total RNA using the High Capacity cDNA Archive Kit. PCR cycling was performed on a 7900HT Sequence Detection System in 25 µl SYBR Green PCR Master Mix using 25 ng of reverse transcribed RNA. Target genes were tested for using Assay-on-Demand Gene Expression products. The GUSB gene was used as an endogenous reference to control for expression independent sample to sample variability. Relative expression was determined from the obtained Ct values and the $2^{-\Delta\Delta Ct}$ method.[13] All equipment and reagents were purchased from Applied Biosystems (Foster City, California, USA) and used according to their protocols. Six genes (ALDOB, IL2RB, PDE7B, TM4SF4, TXN, and TYK2), together with the GUSB reference, were tested in duplicate on pooled cDNA from M0 normal controls (n = 16) and MIII coeliac patients (n = 15). The pooling method was used only after we had assessed that the mean of the data obtained with ALDOB tested on individual samples was similar to that of the pooled samples (table 2). This pooling method was in agreement with our experiences with various other genes not directly related to this study.

Table 2 Relative levels of average gene expression levels in Marsh III coeliac patients versus normal controls, as determined by microarray hybridisation and validation by real time reverse transcription-polymerase chain reaction

Gene	Microarray	RT-PCR
ALDOB*	0.51	0.23
ALDOB†	0.51	0.28
IL2RB2	1.36	1.28
PDE7B	0.64	0.94
TM4SF4	0.50	0.12
TXN	1.26	1.33
TYK2	1.30	0.99

Real time RT-PCR data for ALDOB are indicated as the mean of individual samples* and as pooled samples† of normal controls (n = 16) and Marsh III coeliac patients (n = 15). The other genes were tested only on pools.

RESULTS

Identification of genes differentially expressed in MIII biopsies versus M0 biopsies

In order to identify genes involved in the pathogenesis and therefore in the inflammatory and immune response evoked by gluten and leading to the tissue destruction observed in the duodenal biopsies of coeliac disease patients, the expression profiles of 15 MIII biopsies from patients were compared with seven M0 biopsies from control individuals. All biopsy RNA samples from MIII coeliac disease patients, including four refractory coeliac disease type I samples, were hybridised onto cDNA microarrays. A Welch *t* test with a threshold p value of <0.001 revealed no differentially expressed genes when comparing MIII refractory coeliac disease type I biopsies with MIII coeliac disease biopsies (data not shown). As these two groups could not be distinguished based on their expression levels, they were treated as a single group based on their histological characteristics.

From all 19 200 genes present on the arrays, 10 674 had sufficient data for comparison (that is, genes with a quality score >0.01). A Welch *t* test with a threshold of p<0.001 was applied to the data set and 109 genes were identified to be differentially expressed between MIII and M0 biopsy samples. Approximately 11 genes were expected by chance alone. A distance measurement was used to define the similarity of the expression profiles for both the 22 samples and the 109 genes, and was depicted in a two dimensional hierarchical dendrogram (fig 1A). The hierarchical cluster analysis revealed two branches clearly separating the M0 and MIII biopsies, with the exception of control No 19 that clustered with the group of patient samples.

Of these 109 genes, 76 (69.7%) had an increased level of expression in coeliac disease patients versus controls, and 33 genes (30.3%) had a decreased level of expression in coeliac disease patients versus controls. The molecular function was known or could be predicted for 46 of these 109 genes (table 3).

Some of the differentially expressed genes result from the histological changes but a significant number of them result from the inflammatory response seen in MIII lesions of coeliac disease patients. The results from these experiments are consistent with the widely accepted T helper (Th)1 response.[14] Although this observation in itself adds little new insight, it does replicate previous independent findings. Hence these observations support our results and thereby validate microarray technology as a useful tool in providing new insight into the pathogenesis of coeliac disease.

Although the key cytokines interferon (IFN)-γ and tumour necrosis factor α were not present on our slides, real time RT-PCR of the *IFNG* gene showed extremely high levels of expression of ~30-fold in MIII biopsies versus M0 biopsies. Stimulation of the interleukin (IL)-2 signalling pathway was further suggested by increased expression of the IL-2 receptor beta (*IL-2RB*) gene which was validated by real-time RT-PCR (table 2), a member of the RAS oncogene family (*RAB1B*), chromosome 20 open reading frame 64 gene (*C20orf64*), and the RAP1, GTP-GDP dissociation stimulator 1 gene (*RAP1GDS1*). Furthermore, the nuclear factor κB pathway seems to be induced, as suggested by upregulation

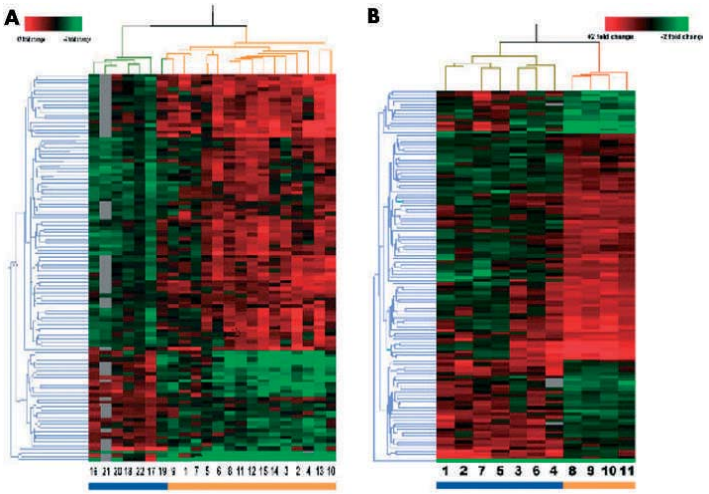


Figure 1 (A) Hierarchical clustering dendrogram of duodenal genes from Marsh III (MIII) and Marsh 0 (M0) biopsies. Clustering of 109 genes across 22 samples clusters the seven control samples (blue bar) separately from the 15 coeliac disease patients (orange bar). Each column represents a coeliac disease (MIII) or a control (M0) sample and each row represents an individual gene. For each gene, a green signal represents underexpression, black signals denote similarly expressed genes, a red signal represents overexpressed genes, and grey signals denote missing data. Information on these genes can be found in table 3 and on the website http://humgen.med.uu.nl/publications/CD/Microarrays/Diosdado_1.html. (B) Hierarchical clustering dendrogram of duodenal genes from MIII biopsies with or without exposures to gluten. Clustering of 120 genes across 11 samples clusters the seven samples of coeliac disease patients following a gluten containing diet (blue bar) separately from the four coeliac disease patients on a gluten free diet (orange bar). Each column represents a sample and each row represents an individual gene. For each gene, a green signal represents underexpression, black signals denote similarly expressed genes, a red signal represents overexpressed genes, and grey signals denote missing data. Information on these genes can be found in table 3 and on the website http://humgen.med.uu.nl/publications/CD/Microarrays/Diosdado_1.html.

of the thioredoxin (*TXN*) gene and the lymphocyte cytosolic protein 1 (*LCP1*) gene. Genes involved in complementary functions, such as complement component 8 β polypeptide (*C8B*) and the H factor 1 (*HF1*), were downregulated, suggesting depletion of complement components.

Upregulation of both the *TXN* and the macrophage scavenger receptor 1 (*MSR1*) genes in MIII versus M0 provides evidence for the presence of active macrophages at the lesion site. MSR1 is expressed by antigen presenting cells (APC) and mediates activation of T cells and promotes adhesion of activated B cells.[15 16] It has also been suggested that MSR class A is involved in the breakdown of the T cell self tolerance in mice.[17 18] TRX enhances the immune response by facilitating both a microenvironment for APCs and for T cell interaction. It has been further proposed that TRX acts as coadjuvant in cytokine mediated lymphocyte proliferation between the APC which presents the gluten to the CD4+ T cells.[19] Upregulation of *TRX* was validated by real time RT-PCR (table 2).

Many genes coding for proteins involved in lipid metabolism and cholesterol homeostasis, such as ATP binding cassette, subfamily A, member 7 (*ABCA7*), apolipoprotein A-II (*APOA2*), and diaphorase (NADH) (*DIA1*), were upregulated in MIII biopsies whereas hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid 1 (*HSD3B1*) was downregulated. Expression studies have shown that γδ T cells differentially express a large number of genes involved in lipid and cholesterol homeostasis, implying that such cells may be involved in these mechanisms.[20] One of the histological characteristics of MIII coeliac disease biopsies is the presence of high numbers of intraepithelial γδ T cells.[3] Hence these results may reflect the increased number of γδ T cells in the intestines of coeliac disease patients. Alternatively, increased expression of lipid and cholesterol genes may point to increased biosynthesis of lipid bilayers. In both cases, upregulation of these genes may be due to an increase in the number of cells and not to upregulation "per cell" of these genes.

Downregulation of the transmembrane 4 superfamily member 4 gene (TM4SF4), which by real time RT-PCR turned out to be down-regulated eightfold (table 2), and upregulation of the retinol binding protein 4 gene (*RBP4*), are both consistent with increased proliferation of cells in MIII biopsies. Also, upregulation of annexin A6 (*ANXA6*), synaptopodin (*KIAA1029*), and the solute carrier family 9 isoform 3 regulatory factor 1 (*SLC9A3R1*) genes suggests an increased number of cells in early differentiation.

In summary, the differentially expressed genes point towards an increased number of γδ T cells and macrophages at the lesion site, a Th1 response, and increased cell proliferation.

Differential gene expression in MIII biopsies in response to gluten withdrawal

To identify genes specifically involved in the response to the environmental trigger gluten, MIII biopsies were subsequently divided into two groups. Four MIII biopsies from coeliac disease patients who were on a gluten free diet for one year were compared with seven MIII biopsies from patients who were following a gluten containing diet at the time of biopsy. The four biopsies of the refractory coeliac disease type I patients were excluded from this analysis as these patients do not show clinical and histological improvement on gluten withdrawal, and so different molecular events may be occurring in their biopsies.

Table 3 Annotated genes differentially expressed in Marsh III (MIII) versus Marsh 0 (M0), grouped according to their function*

Gene symbol	Gene name	LocusLink ID	Chromosome location	Ratio MIII/M0†
Immune response related				
C8B	Complement component 8, beta polypeptide	732	1p32.1	0.67
HF1	H factor 1 (complement)	3075	1q31.3	0.80
MSR1	Macrophage scavenger receptor 1	4481	8p22	1.25
TXN	Thioredoxin	7295	9q31.3	1.26
IL2RB	Interleukin 2 receptor β	3560	22q13.1	1.36
Chemokines, cytokines, and growth factors				
BTN2A1‡	Butyrophilin, BTF1 precursor	11120	6p22.2	1.18
Inflammatory mediators				
PRG1	Proteoglycan 1, secretory granule	5552	10q22.1	1.21
Cancer related genes				
DKFZP586A011	Cervical cancer 1 proto oncogene	25875	12q13.13	0.72
C20orf64	Chromosome 20 open reading frame 64	112858	20q13.12	1.30
Structural related genes				
TM4SF4	Transmembrane 4 superfamily member 4	7104	3q25.1	0.50
IGFBP7	Insulin-like growth factor binding protein 7	3490	4q12	0.73
EPB42	Erythrocyte membrane protein band 4.2	2038	15q15.2	0.79
BTN2A1‡	Butyrophilin, BTF1 precursor	11120	6p22.2	1.18
SLC9A3R1	Solute carrier family 9, isoform 3 regulatory factor 1	9368	17q25.1	1.20
RBP4	Retinol binding protein 4, interstitial	5950	10q23.33	1.24
KIAA1029	Synaptopodin	11346	5q33.1	1.43
Signal transduction				
PPP3CB	Calcineurin A beta	5532	10q22.2	0.76
RAP1GDS1	RAP1, GTP-GDP dissociation stimulator 1	5910	4q23	1.22
RAB1B	RAB1B, member RAS oncogene family	81876	11q13.3	1.23
ANXA6	Annexin A6	309	5q33.1	1.24
LCP1	Lymphocyte cytosolic protein 1	3936	13q14.13	1.30
TYK2	Tyrosine kinase 2	7297	19p13.2	1.30
JFC1	NADPH oxidase related, C2 domain containing protein	84958	1p35.2	1.34
Metabolic pathways and ion transport mediators				
ALDOB	Aldolase B, fructose bisphosphate	229	9q31.1	0.51
PDE7B	Phosphodiesterase 7B	27115	6q23.3	0.64
FLJ12899	Pantothenate kinase 3	79646	5q34	0.71
HSD3B1	Hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid-1	3283	1p12	0.82
ABCA7	ATP binding cassette, subfamily A, member 7	10347	19p13.3	1.18
APOA2	Apolipoprotein A-II	336	1q23.3	1.21
SRI	Sorcin	6717	7q21.12	1.23
GYG2	Glycogenin 2	8908	Xp22.33	1.27
SMUG1	Single strand selective monofunctional uracil DNA glycosylase	23583	12q13.13	1.29
MRPS22	Mitochondrial ribosomal protein S22	56945	3q23	1.29
KCNQ3	Potassium voltage gated channel, KQT-like subfamily, member 3	3786	8q24.22	1.30
ATP1A2	ATPase, Na+/K+ transporting, alpha 2 (+) polypeptide	477	1q23.2	1.30
DIA1	Diaphorase (NADH)	1727	22q13.2	1.30
Cell cycle regulators and transcription factors				
HBP1	HMG-box containing protein 1	26959	7q22.3	0.83
KHDRBS1	Sam 68	10657	1p35.1	1.11
NAP1L4	Nucleosome assembly protein 1-like 4	4676	11p15.4	1.27
C20orf64	Chromosome 20 open reading frame 64	112858	20q13.12	1.30
C21orf66	Chromosome 21 open reading frame 66	94104	21q22.11	1.38
Cell-cell signalling				
ECM2	Extracellular matrix protein 2	1842	9q22.31	1.24
RIMS1	Regulating synaptic membrane exocytosis 1	22999	6q13	1.29
GRM3	Glutamate receptor, metabotropic 3	2913	7q21.12	1.37
Protein synthesis				
EIF4EBP3	Eukaryotic translation initiation factor 4E binding protein 3	8637	5q31.3	0.72
EIF2B3	Eukaryotic translation initiation factor 2B, subunit 3	8891	1p34.1	1.27

*All of the genes can be found at http://humgen.med.uu.nl/publications/CD/Microarrays/Diosdado_1.html.
†Ratio of the expression level for each individual gene when comparing MIII coeliac disease biopsies with M0 biopsies. Genes with values <1 are downregulated; genes with values >1 are upregulated.
‡Genes located in more than one functional group.

A Welch *t* test was applied to 11 938 genes that had sufficient data for comparison; 120 genes showed differential expression at a threshold p value <0.005 (table 4). We are aware of the limitations of this small number of samples. This experimental design is therefore less robust than the MIII versus M0 comparison as the number of genes to be expected by change alone would be 60. Although the data should be interpreted with care, the general picture that emerges from the data may give inroads into the effect of gluten on the intestine. A two dimensional hierarchical cluster showed a dendrogram tree (fig 1B) in which the four patients on a gluten free diet clearly clustered separately from the seven patients on a gluten containing diet.

Table 4 Annotated genes differentially expressed in Marsh III (MIII) biopsies with or without gluten, grouped according to functional pathways*

Gene symbol	Gene name	LocusLink ID	Chromosome location	Ratio gluten−/gluten+†
HLA and immune function genes				
C5R1	Complement component 5 receptor 1	728	19q13.32	1.14
CD79B	CD79B antigen	974	17q23.3	1.14
MAP4K4	Mitogen activated protein kinase kinase kinase kinase 4	9448	2q11.2	1.31
ITGA4	Integrin, alpha 4	3676	2q31.3	1.39
Chemokines, cytokines, and growth factors				
None				
Inflammatory mediators				
PDE4D	Phosphodiesterase 4D, cAMP specific	5144	5q11.2	0.67
AUH‡	AU specific RNA binding protein	549	9q22.31	0.86
Cancer related genes				
DEPC-1	Prostate cancer antigen-1	221120	11p12	0.79
M17S2	Membrane component, chromosome 17, surface marker 2	4077	17q21.31	1.24
RABL2B	RAB-like protein 2B	11158	22q13.33	1.35
Structure related genes				
ADAMTS9	A disintegrin-like and metalloprotease with thrombospondin type 1 motif, 9	56999	3p14.1	0.74
ADRM1	Adhesion regulating molecule 1	11047	20q13.33	0.82
KL	Klotho	9365	13q13.1	1.11
FN1	Fibronectin 1	2335	2q35	1.30
Metabolic pathways and ion transport mediators				
PREP	Prolyl endopeptidase	5550	6q21	0.56
PLCB4	Phospholipase C, beta 4	5332	20p12.2	0.79
AUH‡	AU specific RNA binding protein	549	9q22.31	0.86
RYR3	Ryanodine receptor 3	6263	15q14	1.22
IDH3A	Isocitrate dehydrogenase 3, alpha subunit	3419	15q25.1	1.21
BG1	Lipidosin	23205	15q23	1.21
SLC25A4	Adenine nucleotide translocator 1	291	4q35.1	1.25
CPT1A	Carnitine palmitoyltransferase I, liver	1374	11q13.3	1.31
Cell cycle regulators and transcription factors				
LOC51174	Delta-tubulin	51174	17q23.2	0.54
KIF5B	Kinesin family member 5B	3799	10p11.22	0.76
CCNG1	cyclin G1	900	5q34	0.77
NT5C2	5′-nucleotidase, cytosolic II	22978	10q24.32	0.83
ZNF26	Zinc finger protein 26	7574	12q24.33	1.19
POLR2C	Polymerase II, RNA subunit C	5432	16q13	1.19
GTF2H1	General transcription factor IIH, polypeptide 1	2965	11p15.1	1.24
MSI2	Musashi homologue 2 (Drosophila)	124540	17q23.2	1.28
ZNF317	Zinc finger protein 317	57693	19p13.2	1.32
CDC25A	Cell division cycle 25A	993	3p21.31	1.46
PPP2R3A	Protein phosphatase 2 (formerly 2A), regulatory subunit B′, alpha	5523	3q22.3	1.49
POLD3	DNA polymerase delta subunit 3	10714	11q13.4	1.58
Signal transduction				
SYPL	Synaptophysin-like protein	6856	7q22.3	0.79
WAC	WW domain containing adapter with a coiled-coil region	51322	10p12.1	0.83
HNT	Neurotrimin	50863		0.86
RGS16	Regulator of G protein signalling 16	6004	1q25.3	1.23
PSCD3	Pleckstrin homology, Sec7, and coiled/coil domains 3	9265	7p22.1	1.33
CSPG6	Chondroitin sulphate proteoglycan 6 (bamacan)	9126	10q25.2	1.36
Others				
	Pro-neuregulin-3 precursor		10q23.1	0.69
HSPA6	Heat shock 70 kDa protein 6	3310	1q23.3	0.76
SARM	Sterile alpha and HEAT/Armadillo motif protein, orthologue of Drosophila	23098	17q11.2	1.26
AMSH	Associated molecule with the SH3 domain of STAM	10617	2p13.1	1.29
PERQ1	Postmeiotic segregation increased 2-like 12	64599	7q22.1	1.59
HTF9C	HpaII tiny fragments locus 9C	27037	22q11.21	1.61

*All of the genes can be found at http://humgen.med.uu.nl/publications/CD/Microarrays/Diosdado_1.html.
†Ratio of the expression level for each individual gene between biopsies from MIII coeliac disease patients on a gluten free diet (gluten−) and MIII coeliac disease patients following a gluten containing diet (gluten+). Genes with values <1 are downregulated; genes with values >1 are upregulated.
‡Genes located in more than one functional group.

Of these 120 genes, 46 (38.3%) were downregulated and 75 (62.5%) were upregulated in patients on gluten versus coeliac disease patients refraining from gluten. Functional annotation was available for 42 of the 120 genes. Interestingly, the majority of the differentially expressed genes suggest transcription is enhanced and mitotic activity is suppressed in MIII coeliac disease patients who refrain from gluten.

Several of the genes were related to cell division and cell cycle processes. The cyclin G1 (*CCNG1*) gene was downregulated, indicating that less cells go from the G1 to S phase. Both the delta-tubulin (*LOC51174*) and kinesin family member 5B (*KIF5B*) genes were also downregulated. These two genes play a role in spindle formation. In addition, genes related to transcription such as general transcription factor IIH, polypeptide 1 (*GTF2H1*), polymerase II RNA subunit C (*POLR2C*), DNA polymerase delta subunit 3 (*POLD3*), zinc finger protein 26 (*ZNF26*), and the zinc finger protein 317 (*ZNF317*) were all upregulated.

DISCUSSION

Coeliac disease is a chronic inflammatory disorder triggered by dietary gluten that shows major manifestations in the small intestine although other organ systems may also be involved. Coeliac disease is an HLA associated disease and over the past years our understanding of the role of both HLA-DQ2 and -DQ8 molecules and gluten polypeptides has been enhanced tremendously.[2] Nevertheless, a detailed understanding of the molecular events that take place in the intestine of genetically susceptible individuals is lacking, as well as insight into the order in which the events occur and other genetic factors that are known to be important determinants of the disease. To unravel novel aspects of the pathogenesis of coeliac disease, cDNA microarray technology was used to monitor expression of thousands of genes simultaneously. The results from this study confirm earlier studies that coeliac disease is a Th1 mediated disease,[4 21] as evidenced by upregulation of the IL-2 pathway. Interestingly, novel genes are being identified that have not been proposed previously as being important determinants of the pathogenesis of coeliac disease. These genes yield new insights into the molecular processes underlying the flattened mucosa.

Prolyl endopeptidase

The most notable gene differentially expressed in a comparison of coeliac disease patients on a gluten containing diet versus those on a gluten free diet, is prolyl endopeptidase (*PREP*). The *PREP* gene encodes a cytosolic prolyl endopeptidase that efficiently hydrolyses proline rich fragments such as gliadin. *PREP* is upregulated in coeliac disease patients on a gluten containing diet compared with patients adhering to a gluten free diet. Furthermore, its levels of expression stay elevated in controls. Interestingly, this gene was also shown to be upregulated in seven control samples not on a gluten free diet compared with four biopsies from coeliac disease patients on a gluten free diet ($p<0.003$) and, to a lesser extent, when compared with the four refractory coeliac disease samples on a gluten free diet ($p<0.025$). This observation indicates that it is the presence of gluten in the intestine which modulates expression of this gene. It has been hypothesised that prolyl endopeptidase cleaves gliadin in small fragments in the intestine.[22] We hypothesise that, in coeliac disease patients, the activity of this molecule may be impaired and therefore it does not cut gliadin into sufficiently small fragments to create motifs long enough to be recognised by reactive T cells.[22] It has been proposed that a bacterial prolyl endopeptidase from *Flavobacterium memningoseticum* may be a target for future treatment of coeliac disease patients.[23] Further research should be done in order to elucidate the mechanisms controlling activation and expression of this protein in the intestines of normal and coeliac disease patients.

Molecular mechanisms controlling hyperplasia and villous atrophy in MIII biopsies

Enterocytes, one of the main components of the villous, arises from pluripotent stem cells located in the base of the intestinal crypts. When these pluripotent stem cells stop dividing, they complete their programme of differentiation and start migrating towards the apical part of the villous. Non-dividing epithelial cells located in the crypts of the small intestine that are ready to migrate and differentiate express the TM4SF4 gene.[24] In vitro studies have suggested that TM4SF4 plays a crucial role in regulation of proliferation and differentiation along the crypts by inhibiting proliferation of cells at the boundary of the crypt and villous. We observed an eightfold decrease in the level of expression of the TM4SF4 gene in MIII biopsies of coeliac disease patients versus M0 biopsies of controls, implying a block in early differentiation and failure to complete villous maintenance. Furthermore, increased expression of the *RBP4* gene and significant upregulation ($p<0.01$) of the Eph related receptor tyrosine kinase B3 (*EphB3*) gene in MIII versus M0 biopsies was observed. EphB3 has recently been shown to be tightly regulated by the Wnt cascade and it has been suggested that this gene may play a decisive role in controlling not only proliferation and differentiation but also migration and location of cells along the crypts.[25]

Given these results, we propose that stem cells of the small intestinal crypts start proliferating but do not receive a signal to start differentiation, leading to the formation of undifferentiated hyperplastic crypts and subsequently villous atrophy. This theory is further substantiated by results from the second experiment in which MIII lesions with or without exposure to gluten were compared. The presence of gluten leads to increased mitotic activity, suggesting many proliferating cells. As soon as gluten is withdrawn from the diet, mitotic activity diminishes, suggesting that cells stop proliferating and, presumably, start differentiating. Although the histology still shows a MIII lesion, this event might actually be the first step towards normalisation of the intestine. We further propose that the separate processes of crypt hyperplasia and villous atrophy are molecularly intertwined and progression through different histological stages may be regulated by genes implicated in the control of cell proliferation and differentiation, such as TM4SF4.

So far, the early pathogenic events leading to villous atrophy have been suggested to be direct effects of cytokines, such as IFN-γ,[21] or hypoxia.[3] In addition, using an ex vivo model of fetal gut, Salmela and colleagues recently proposed that increased levels of metalloproteinases are important molecules for tissue remodelling and mucosal degradation in inflammatory bowel disease and coeliac disease.[26] However, none of the 10 metalloproteinase family members, nor two of its inhibitors that were present on our slides, showed a significant change in expression level in MIII biopsies of coeliac disease patients (data not shown), implying that these molecules may not play a direct role in the tissue damage in vivo. Until now, all studies have proposed that villous atrophy is caused by *destruction* of enterocytes and, consequently, loss of structure of the villous. We propose that villous atrophy is not due to destruction of the villous but rather to the failure of crypt cells to differentiate into fully differentiated villous cells. As the cells seem to be capable of continuing proliferation, this might explain crypt hyperplasia.

We have successfully demonstrated the use of microarrays as a general approach to studying complex human diseases. Expression profiles of the selected biopsies propose known genes that participate in the pathogenesis of coeliac disease as well as new pathways that may play a potential role in disease initiation and progression. Our results suggest that

crypt hyperplasia may be explained by distortion in the ratio between cell proliferation and cell differentiation of cells composing the crypt-villi units. Further investigations of the identified genes are required and will hopefully advance our understanding of the molecular mechanisms underlying coeliac disease. Finally, the finding that prolyl endopeptidase is differentially expressed in coeliac disease patients may provide new leads for therapy.

ACKNOWLEDGEMENTS

The authors thank Dr Frits Koning and Darius Stepniak for helpful discussions, Daniel Chan for helping with the illustration, Justin Fransen for the real time RT-PCR, Jackie Senior for critically reading the manuscript, and the coeliac disease families for their interest and collaboration. This work was financially supported by the Dutch Digestive Disease Foundation (MLDS) grant WS00-13, the Netherlands Organisation for Medical and Health Research (NWO) grant 902-22-094, the KNAW Ter Meulen Fonds, and a Fulbright fellowship to CW.

Authors' affiliations

B Diosdado, M C Wapenaar, L Franke, C Wijmenga, Complex Genetics Group, Department of Biomedical Genetics, University Medical Centre, Utrecht, the Netherlands
K J Duran, Complex Genetics Group, and Genomics Laboratory, Department of Biomedical Genetics, University Medical Centre, Utrecht, the Netherlands
M J Goerres, Department of Gastroenterology and Hepatology, Rijnstate Hospital, Arnhem, the Netherlands
M Hadithi, C J J Mulder, Department of Gastroenterology and Hepatology, Rijnstate Hospital, Arnhem, the Netherlands, and Department of Gastroenterology, Free University Medical Centre, Amsterdam, the Netherlands
J B A Crusius, Laboratory of Gastrointestinal Immunogenetics, Free University Medical Centre, Amsterdam, the Netherlands
J W R Meijer, Department of Pathology, Rijnstate Hospital, Arnhem, the Netherlands
D J Duggan*, National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, MD, USA
F C P Holstege, Genomics Laboratory, Department of Physiological Chemistry, University Medical Centre, Utrecht, the Netherlands

*Present address: TGen, Translational Genomics Research Institute, Phoenix, AZ, USA

REFERENCES

1 **Vader W**, Kooy Y, Van Veelen P, *et al*. The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* 2002;**122**:1729–37.
2 **Sollid LM**. Molecular basis of celiac disease. *Annu Rev Immunol* 2000;**18**:53–81.
3 **Marsh MN**. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 1992;**102**:330–54.
4 **Sollid LM**. Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* 2002;**2**:647–55.
5 **Schuppan D**. Current concepts of celiac disease pathogenesis. *Gastroenterology* 2000;**119**:234–42.
6 **Lockhart DJ**, Winzeler EA. Genomics, gene expression and DNA arrays. *Nature* 2000;**405**:827–36.
7 **Mulder C**, Rostami K, Marsh MN. When is a coeliac a coeliac? *Gut* 1998;**42**:594.
8 **Mulder CJ**, Wahab PJ, Moshaver B, *et al*. Refractory coeliac disease: a window between coeliac disease and enteropathy associated T cell lymphoma. *Scand J Gastroenterol Suppl* 2000;**232**:32–7.
9 **Van de Peppel JPK**, van Bakel H, Radonjic M, *et al*. Monitoring global mRNA changes with externally controlled microarray experiments. *EMBO* 2002;**4**:387–93.
10 **Brazma A**, Hingamp P, Quackenbush J, *et al*. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat Genet* 2001;**29**:365–71.
11 **Chen Y**, Kamat V, Dougherty ER, *et al*. Ratio statistics of gene expression levels and applications to microarray data analysis. *Bioinformatics* 2002;**18**:1207–15.
12 **Chen Y**, Kamat V, Dougherty ER, *et al*. Ratio statistics of gene expression levels and applications to microarray data analysis. *Bioinformatics* 2002;**18**:1207–15.
13 **Livak KJ**, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)). *Methods Method* 2001;**25**:402–8.
14 **Nilsen EM**, Jahnsen FL, Lundin KE, *et al*. Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease. *Gastroenterology* 1998;**115**:551–63.
15 **Nicoletti A**, Caligiuri G, Tornberg I, *et al*. The macrophage scavenger receptor type A directs modified proteins to antigen presentation. *Eur J Immunol* 1999;**29**:512–21.
16 **Yokota T**, Ehlin-Henriksson B, Hansson GK. Scavenger receptors mediate adhesion of activated B lymphocytes. *Exp Cell Res* 1998;**239**:16–22.
17 **Choudhury A**, Mukherjee P, Basu SK, *et al*. Disruption of T cell tolerance to self-immunoglobulin causes polyclonal B cell stimulation followed by inactivation of responding autoreactive T cells. *J Immunol* 2000;**164**:1713–21.
18 **Abraham R**, Choudhury A, Basu SK, *et al*. Disruption of T cell tolerance by directing a self antigen to macrophage-specific scavenger receptors. *J Immunol* 1997;**158**:4029–35.
19 **Angelini G**, Gardella S, Ardy M, *et al*. Antigen-presenting dendritic cells provide the reducing extracellular microenvironment required for T lymphocyte activation. *Proc Natl Acad Sci U S A* 2002;**99**:1491–6.
20 **Fahrer AM**, Konigshofer Y, Kerr EM, *et al*. Attributes of gammadelta intraepithelial lymphocytes as suggested by their transcriptional profile. *Proc Natl Acad Sci U S A* 2001;**98**:10261–6.
21 **Nilsen EM**, Lundin KE, Krajci P, *et al*. Gluten specific, HLA-DQ restricted T cells from coeliac mucosa produce cytokines with Th1 or Th0 profile dominated by interferon gamma. *Gut* 1995;**37**:766–76.
22 **Hausch F**, Shan L, Santiago NA, *et al*. Intestinal digestive resistance of immunodominant gliadin peptides. *Am J Physiol Gastrointest Liver Physiol* 2002;**283**:G996–1003.
23 **Shan L**, Molberg O, Parrot I, *et al*. Structural basis for gluten intolerance in celiac sprue. *Science* 2002;**297**:2275–9.
24 **Wice BM**, Gordon JI. A tetraspan membrane glycoprotein produced in the human intestinal epithelium and liver that can regulate cell density-dependent proliferation. *J Biol Chem* 1995;**270**:21907–18.
25 **Batlle E**, Henderson JT, Beghtel H, *et al*. Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/EphrinB. *Cell* 2002;**111**:251–63.
26 **Salmela MT**, MacDonald TT, Black D, *et al*. Upregulation of matrix metalloproteinases in a model of T cell mediated tissue injury in the gut: analysis by gene array and in situ hybridisation. *Gut* 2002;**51**:540–7.

Chapter 4

A genomics view on coeliac disease points to enhanced neutrophil recruitment and barrier impairment

Submitted

A genomics view on coeliac disease points to enhanced neutrophil recruitment and barrier impairment

Begoña Diosdado[1], Harm van Bakel[1], Eric Strengman[1], Lude Franke[1], Erica van Oort[1], Chris J Mulder[2], Cisca Wijmenga[1], and Martin C Wapenaar[1]

[1] *Complex Genetics Section, Department of Medical Genetics, University Medical Center, Utrecht;*
[2] *Department of Gastroenterology, VU Medical Center, Amsterdam, The Netherlands*

***Background and Aim*: Coeliac disease is an enteropathy featuring villous atrophy, crypt hyperplasia, and lymphocytosis. Tissue remodeling is driven by an inflammatory reaction to gluten in genetically susceptible individuals. The adaptive Th1 pathway is considered the major immune response but recent evidence indicated involvement of innate immunity as well. To assess the contribution of either immune response we performed global gene expression profiling of the regenerating mucosa.**

***Methods*: Microarray hybridizations were performed with biopsy samples from 13 untreated patients, 31 patients on a gluten-free diet in various stages of remission, and 21 controls. Supplementary data were generated using low density array and conventional quantitative reverse-transcription PCR, and immunohistochemistry.**

***Results*: 108 differentially expressed immune-related genes were identified (44 innate, 43 adaptive, nine both innate/adaptive, and six immuno-regulatory). Expression levels plotted as a function of the tissue regeneration sequence showed a gradual change as opposed to the discrete histological transitions. In addition to details provided on the adaptive and innate immune pathways utilized, we observed a chronic recruitment of activated neutrophils. Neutrophil involvement was unabated in otherwise completely normalized remission patients.**

***Conclusion*: We observed a comparable contribution of both the innate and adaptive immune response in coeliac disease pathogenesis. The discrepancy between the histological classification and the observed incremental change in immune-gene expression may have consequences for current diagnostic inclusion criteria. Enhanced neutrophil infiltration in both active and remission patients points to a genetic impairment of the intestinal barrier that may contribute to the cause rather than the consequence of coeliac disease.**

INTRODUCTION

Coeliac disease (CD) is an inflammatory condition of the intestine with the unique quality for a complex genetic disorder that its features can be provoked or suppressed by manipulating its principal environmental trigger: dietary gluten. This makes the disease particularly amenable to an experimental approach to study its molecular pathology. The flattened intestinal mucosa of patients on a gluten-free diet goes through a series of distinct transformations characterized by successive recovery from villous atrophy, crypt hyperplasia, and lymphocytosis. These stages were originally described by Marsh and are referred to as MIII to M0.[1] We presume that this sequence of events, together with the underlying molecular processes, occurs in reverse order during the gluten-induced pathogenesis. We reported an earlier gene expression study in which we showed that this inflammation-evoked tissue restructuring was marked by enhanced proliferation and reduced differentiation of cells in the coeliac mucosa.[2] The activation of gluten-restricted T cells through HLA-DQ2 antigen-presentation and the resulting Th1 response is considered to be the major force driving the pathology.[3] This is further underscored by the principal risk contributed by genes encoding for this HLA-type.[4] Recently, the discussion on the

gluten-induced inflammation has widened to include innate immunity as well.[5] However, we need more details on the inflammatory response to fully appreciate the contribution and interactions of both the innate and adaptive immunity in CD. In this study we have addressed this issue by determining differential expression of immune-related genes using microarray technology and low density array quantitative RT-PCR. By analyzing biopsy samples during various stages of recovery, we also obtained insight into the molecular dynamics that accompany this mucosal restructuring. These observations will further fuel the debate on the diagnostic inclusion criteria for CD. We found compelling evidence for the role of neutrophils in the coeliac lesion, a leukocyte that has so far been ignored in studies of the pathogenesis. Based on these observations, and together with our recent publication on a new coeliac susceptibility gene,[6] we propose a model of CD pathogenesis that includes elements of intestinal permeability, neutrophil recruitment, and Th1 activation.

MATERIALS AND METHODS

Patients and biopsy samples

Details on the patients used in this study, the Marsh classification of their duodenal biopsy samples, and the expression assays applied, are summarized in supplementary table 1. In brief, duodenal biopsies from 49 coeliac patients which were diagnosed according to the modified ESPGHAN criteria were included in this study.[7] Sixteen of them were untreated at the time of diagnosis and their biopsies showed Marsh III lesions (villous atrophy, crypt hyperplasia, and lymphocytosis). The remaining 31 patients were diagnosed previously with Marsh III lesions and were on a gluten-free diet at the time of biopsy sampling for a routine follow-up check. All treated patients showed clinical improvement and their mucosal biopsies displayed various stages of remodeling towards normalization. Each stage was classified by the presence or absence of villous atrophy (MIII), crypt hyperplasia (MIII-MII), and lymphocytosis (MIII-MI) according to the Marsh nomenclature.[1] Patients in complete remission (M0) were histologically comparable to controls. Controls were samples obtained from 26 individuals who were subjected to endoscopic examination for various reasons, and displayed a normal duodenal histology and no serological indication of CD. Written informed consent was obtained from all participants and the study was approved by the Medical Ethics Committee of the University Medical Center Utrecht.

Expression studies

RNA isolation and microarray hybridization

Two to three duodenal biopsies were obtained by endoscopy from each individual and snap frozen and stored in liquid nitrogen. A surgical specimen from the small bowel was used as a reference sample for all microarray hybridizations. Frozen biopsies were homogenized and total RNA was isolated as previously described.[8] Total RNA from the reference and the biopsy samples was amplified to cRNA. RNA amplification was performed with 1 µg of total RNA by *in vitro* transcription with T7 RNA-polymerase and 5-(3-aminoallyl)-UTP following the manufacture's instructions (Ambion Inc. TX, USA). Next, 1000 ng of each cRNA sample was used to incorporate either Cy3 or Cy5 label (Amersham Pharmacia Biotech, NJ, USA) as previously described.[2] A pool was made of either Cy3- or Cy5-labeled cRNA from the reference. Samples were used only for subsequent hybridization when the incorporation frequency of labeled nucleotides was between 1.5–2%. Dye-swap experiments were performed for all biopsy samples against the common reference. Hybridizations, 132 in total, were performed on home-made microarray slides containing 21,329 genes spotted as 70-mer oligonucleotides.[9] All samples were hybridized as dye-swap duplicates using 500 ng of either Cy3- or Cy5-labeled cRNA against the reference labeled with the complementary dye. After overnight hybridization at 37°C, slides were washed manually and scanned with the Agilent G256AA DNA microarray scanner (100% laser power, 30% photomultiplier tube). All microarray expression data and protocols were deposited in the MIAME compliance database ArrayExpress.[10]

Microarray data analysis

Quantification of all the signals on the arrays was performed with Imagene v.5.6 software

(BioDiscovery, Inc. CA, USA). Data was normalized per subgrid using the marrayNorm R package v.1.1.3. and its variance was stabilized with the VSN R package v.1.3.2, as described previously.[11] Expression of each gene was calculated as the ratio between biopsy to reference. ANOVA analysis was applied to select differentially expressed genes among the various patient and control groups ($p<0.05$). Additional differentially expressed genes were identified using a standard correlation method that took into account the Marsh recovery sequence (MIII–MII–MI–M0). The ANOVA analysis was performed using the MAANOVA R package v.0.95-3. We used a fixed-effects model that took into account array and dye effects, and only selected genes with a false-discovery rate (FDR) adjusted tabulated p-value of $q<0.05$. The second analysis based on the standard correlation (95%)(GeneSpring package) took into account the expression pattern of the MAANOVA selected genes at the consecutive Marsh stages. We used the GeneSpring package v.6.1 and only genes with at least 1.25-fold change were selected. In order to visualize the data, all signals per gene for all individuals belonging to the same Marsh class were averaged and normalized against the average gene signal of the untreated MIII group. Expression profiles of the genes were visualized with GeneSpring software version v.6.1 (Silicon Genetics, Redwood City, CA, USA). We used TEAM[12] for functional annotation of differentially expressed genes and to select the genes with an immune-related function.

Quantitative Reverse-Transcription PCR

Quantitative RT-PCR was performed with either Pre-Configured Low Density Arrays (LDA) or conventional single qRT-PCR using the ABI 7900HT platform. All equipment, software, and reagents were purchased from Applied Biosystems (Foster City, California, USA) and used according to their protocols. We used the Taqman Low Density Immune Profiling Array (part nr 4342510) with the internal endogenous control GUSB to assess the relative expression of 14 MIII samples normalized to the signal of a pool of eight cDNAs from healthy controls (see supplementary Table 1 for the samples used). Genes were measured in fourfold. Relative expression was determined with SDS2.1 software.

Conventional single qRT-PCR was performed for the CYP4F2 (Hs00426608_m1) and CYP4F3 (Hs00168521_m1) genes, again with the GUSB gene (PARD 4326320E) as endogenous control, as previously described.[8] All samples were measured in duplicate. Relative expression was determined using SDS2.1 software and the 2(-Delta Delta C(T)) method .[13]

Immunohistochemistry

Paraffin-embedded duodenal biopsy sections from two healthy controls, two coeliac patients in complete remission (M0), and two patients with active CD (MIII) were used. De-paraffinized and re-hydrated sections were blocked with peroxidase-blocking buffer (pH 5.8) for 20 minutes at room temperature. High-temperature antigen retrieval treatment was performed for 20 minutes in 0.01 M Tris/EDTA solution (pH 9) for myeloperoxidase primary antibody (Lot 128992, Abcam), or in 0.1M sodium citrate buffer (pH 6) for HLA-E (MEM-E/02, Abcam). PBS-rinsed sections were incubated overnight at 4ºC with either the HLA-E antibody (diluted 1:150) or the myeloperoxidase antibody (diluted 1:200). After rinsing with PBS, the sections were incubated for 30 minutes with Envision kit (DAKO) and then rinsed with phosphate citrate buffer (pH 5.8). Stainings were developed using DAB and counterstained with haematoxylin. All sections were reviewed by a pathologist who was not personally involved in the study.

RESULTS

Expression profiles of immune-related genes in the inflamed intestinal mucosa

CD is an inflammatory condition of the small intestine caused by gluten intolerance. Components of this inflammatory response that specifically evoke the mucosal transformation leading to villous atrophy still have to be identified. A gluten-free diet leads to remission and is accompanied by a sequence of discrete changes in the mucosa. These correspond to the MIII to M0 stages described by Marsh[1] and are characterized by successive recovery from villous atrophy, crypt hyperplasia, and lymphocytosis

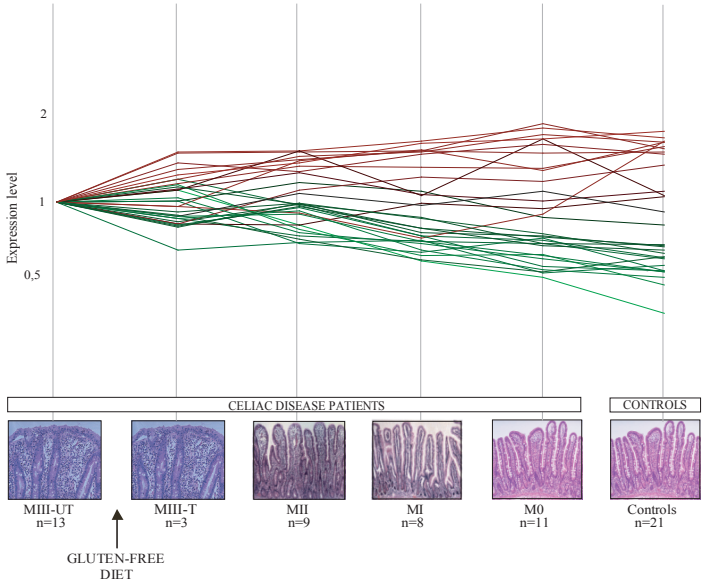


Figure 1. Profiles of differentially expressed immune-related genes during mucosal recovery. The pattern of expression of 27 immune-related genes selected by MAANOVA analysis showed a gradual change as opposed to the discrete histological transitions during remission. The upper panel shows the relative expression of genes that was either enhanced (red) or decreased (green) during remission. Expression was normalized to the values measured in untreated Marsh III patients. The lower panels show the corresponding histology and Marsh classification of the biopsy samples taken before treatment and during remission on a gluten-free diet. Histopathological features include villous atrophy (MIII), crypt hyperplasia (MIII-MII), and lymphocytosis (MIII-MI). Patients in complete remission (M0) are comparable to controls. MIII-UT: Marsh III untreated; MIII-T: Marsh III treated; MII: Marsh II; MI: Marsh I; M0: Marsh 0. Indicated is the number of independent individuals included in each group.

(Fig. 1). We conducted a genomics study, including global gene expression assays using microarray hybridizations, to determine the transcriptional activity of genes related to the immune response during the recovery process. We wanted to address the following questions: 1) which are the immune genes and immunological processes involved in CD; 2) are the discrete histological transitions during remission accompanied by stage-specific gene expression profiles; and if so, 3) which genes display these profiles. We used Gene Ontology (www.geneontology.org) and manual PubMed searches to classify genes as immune response-related. We analyzed the data by applying both MAANOVA statistics[14] and an explorative method using Standard Correlation. Low density array qRT-PCR was performed to supplement data for immune-related genes not represented on the microarrays. Where required, data were added from individual qRT-PCR tests and immunohistochemistry. The microarray hybridizations were performed with duodenal biopsy samples taken from 13 untreated patients (UT-MIII), 31 patients in various stages of remission (MIII–M0), and 21 healthy controls (see supplementary Table 1 for patient details). The data was analyzed with MAANOVA[14] and a FDR cutoff of 0.05 was applied to detect differential gene expression. This yielded 27 immune-related differentially expressed genes, of which 17 showed reduced transcript levels and ten increased levels following the course of tissue normalization (Table 1).

All gene expression profiles showed a gradual increase or decrease from the moment the gluten-free diet was started up to the point of complete recovery (Fig. 1). Apparently the dynamics of the differential gene expression do not reflect the more or less discrete stages of tissue remodeling seen in histological examination. None of the profiles showed an optimum for one of the intermittent stages of mucosal recovery. This gradual change was not limited to the immune-related genes but was observed for all 118 differentially expressed genes comprising various biological processes

(supplementary Fig. 1).

The relatively low number of differentially expressed genes that passed the significance threshold in the MAANOVA may have resulted from the fact that it considers the Marsh stages as independent classes while in reality they follow a fixed sequence during remission. We therefore used an additional approach based on standard correlation (>95%) and the genes identified by MAANOVA. This yielded 64 immune-related genes that showed a similar temporal behavior with at least a 1.25-fold change between MIII and M0 and a profile of gradual, but consistent, up- or down-expression (Table 1). Reference profiles for standard correlation were derived from the MAANOVA analysis: the classic brush border marker dipeptidylpeptidase IV (DDPIV) for enhanced transcription, and a cumulative trend line (derived from 40 genes) for reduced transcription. We also added 38 immune-related differentially expressed genes (MIII versus controls) derived from low density array qRT-PCR experiments (>2-fold change). In total, 108 independent immune-related genes were identified that revealed differential transcript levels during remission in coeliac patients: 78 genes showed reduced levels while 30 genes had higher levels towards mucosal normalization (Table 1). Grouped according to function, 44 genes are involved in innate immunity, 43 genes in adaptive immunity, nine genes in both innate and adaptive immunity, and six genes in immuno-regulatory pathways. This suggests that, based on the number of transcriptionally activated genes, both innate and adaptive immunity contribute comparably to CD pathogenesis.

The innate immune response

Innate immunity constitutes of several non-adaptive defense mechanisms including an epithelial barrier, secreted antimicrobial peptides, chemokines, intraepithelial lymphocytes, NK cells, and phagocytes. On diet-induced tissue normalization, the expression diminished of Paneth cell-derived defensin 6 (*DEFA6*), trefoil factor 1 (*TFF1*), lipocalin 2 (*LCN2*), Ig alpha-1 chain C region (*IGHA1*), the uteroglobin-related protein 2 (*SCGB3A1*), and the IFNG-regulated genes: 2-5A-dependent ribonuclease (*RNAseL*), interferon-induced guanylate-binding protein 1 (*GBP1*), and interferon-induced guanylate-binding protein 5 (*GBP5*). This mirrors the enhanced antimicrobial activity of the intestinal mucosa during the pro-inflammatory phase. On the contrary, the genes liver-expressed antimicrobial peptide 2 (*LEAP-2*), peptide YY, 2 (*PYY2*) and proline-rich acidic protein 1 (*PRAP-1*) were at higher levels during remission.

On mucosal recovery the expression of leukocyte-specific transcript 1 (*NKp30*), killer cell lectin-like receptor subfamily G, member 1 (*KLRG1*), natural killer cells protein 4 (*NK4*), osteoclast inhibitory lectin isoform 1 (*OCIL*), and nucleotide-binding oligomerization domains 27 (*NOD27*) also decreased. This may suggest the involvement of NK and αβ CD8+ T cells receptors. In addition, the non-classical MHC class I genes, *HLA-E*, *HLA-F* and *HLA-G* and some members of the immunoproteasome (*PMSA2*, *PMSA5*, *PMSA7*, and *PMSA8*) showed reduced transcript levels on remission (Table 1). The elevated expression of HLA-E during the pro-inflammatory phase was supported and further specified by immunohistochemistry, revealing an intense granular cytoplasmatic staining of the protein in all MIII epithelial cells, but only moderate staining of enterocytes at the villus tip in normalized M0 patients and controls (Fig. 2A). HLA-E has been implied in immuno-suppression of NK cells through binding to the CD94/NK2G inhibitory receptor complex.[15] Enhanced HLA-E expression may thus reflect the requirement to balance NK cell activation that occurs through binding of its NKG2D receptor to the MICA ligand expressed on stressed enterocytes.[16] Cytokine IL-15 has been proposed as an important mediator of the innate immune response in CD through the activation of NK cells.[17] Contrary to our expectation, the transcriptional level of *IL15* increased during the remission process (Table 1). This further adds to the complexity of IL15 regulation *in vivo* by the monomyeloid and epithelial cells. IL-12 is a pro-inflammatory cytokine involved in NK cell activation in the early steps of the innate immune response. Active IL-12 is a heterodimer formed by two subunits, p40 and p35. Our results showed that transcripts for the p35 subunit are strongly raised in active CD but that levels of the *p40* subunit were below detection (Table 1 and results

Table 1. Differentially expressed immune-related genes during the mucosal restoration in coeliac disease

Gene symbol	Gene name	Fold change	Regulated[a]	Analysis MAANOVA[b]	Analysis SC[c]	Analysis LDA[d]
ADAPTIVE RESPONSE						
Antigen presentation						
HLA-DOA	MHC class II antigen DOA	0,76	IFN-γ		•	
FCGR1A	High affinity immunoglobulin gamma Fc receptor I precursor	0,79	IFNγ		•	
CTSE	Cathepsin E precursor (EC 3.4.23.34)	1,42			•	
ADAMDEC1	ADAM-like, decysin 1	1,78		•	•	
HLADRB1	Major histocompatibility complex, class II, DR beta 1	2,00				•
Cell surface receptor						
ICOS	Inducible T-cell co-stimulator	0.27				•
IL-2RA	Interleukin-2 receptor alpha	0.28				•
CD86	CD86 antigen	0.41				•
TREM2	Triggering receptor expressed on myeloid cells 2	1.49			•	
CCR4	Chemokine (C-C motif) receptor 4	2.70				•
Chemokine and cytokine activity						
IL-12A	Interleukin-12A	0.05				•
CXCL11	Chemokine (C-X-C motif) ligand 11	0.05	IFN-γ			•
IFNG	Interferon gamma	0.10	T-bet		•	•
IL-2	Interleukin-2	0.12				•
CCL19	Chemokine (C-C motif) ligand 19	0.37				•
PPIB	Peptidyl-prolyl cis-trans isomerase B (EC 5.2.1.8)	0.45		•		
CXCL10	Small inducible cytokine B10	0.49	IFN-γ	•	•	•
CCL5	Small inducible cytokine A5	0.63	IFN-γ	•	•	
HABP4	Hyaluronan binding protein 4	0.63			•	
CCL18	Small inducible cytokine A18	0.66		•	•	
CCL4	Small inducible cytokine A4	0.66		•		
CCRL1	C-C chemokine receptor type 11	0.75			•	
MIF[c]	Macrophage migration inhibitory factor	0.80			•	
CXCL14	Small inducible cytokine B14	1.62		•	•	
IL-18	Interleukin-18	3.23				•
Immune regulation						
B7H1	Programmed cell death 1 ligand 1	0.66	IFN-γ		•	
PAG1	Phosphoprotein associated with glycosphingolipid microdomains	1.70		•	•	
Signal transduction						
CD38	CD38 antigen	0.48				•
JAK2	Tyrosine-protein kinase JAK2 (EC 2.7.1.112)	0.75	IFN-γ		•	
PTPN4	Protein tyrosine phosphatase, nonreceptor type 4 (EC 3.1.3.48)	1.06		•		
TCRIM	T-cell receptor interacting molecule	1.81		•		
Transcription factor						
STAT1	Signal transducer and activator of transcription 1-alpha/beta	0.51	IFN-γ	•	•	
TBX21	T-box transcription factor TBX21	0.65	IFN-γ		•	•
XBP1	X box binding protein-1	0.77			•	
IRF8	Interferon consensus sequence binding protein	0.80	IFN-γ		•	
Proteosome modification						
PSME2*	Proteasome activator complex subunit 2	0.67			•	
PSMA7*	Proteasome subunit alpha type 7 (EC 3.4.25.1)	0.70			•	
PSMB8*	Proteasome subunit beta type 8 precursor (EC 3.4.25.1)	0.70				
PSMA5*	Proteasome subunit alpha type 5 (EC 3.4.25.1)	0.80			•	
Miscellaneous						
IFI27	Interferon-alpha induced 11.5 kDa protein (p27)	0.34	IFN-α/β	•	•	
ZAP70*	Tyrosine-protein kinase ZAP-70 (EC 2.7.1.112)	0.64			•	
HMGB2	High mobility group protein 2	0.73			•	
KPNA2*	Importin alpha-2 subunit	0.73			•	

TXN*	Thioredoxin	0.74			•	
APS	Adaptor prot. with pleckstrin and src homology 2 domains	0.76			•	
IFITM1*	Interferon-induced transmembrane protein 1	0.76	IFN-γ		•	
CAMK2G	Calmodulin-dependent protein kinase type II gamma chain	0.77	IFN-γ		•	
HMGB3	High mobility group protein 3	0.78			•	
IFI35	Interferon-induced 35 kDa protein	0.79	IFN-γ		•	
RNF128	ring finger protein 128 isoform 2	1.50			•	
EDN1	Endothelin 1	2.56				•
ECE	Endothelin converting enzyme 1	6.25				•
INNATE RESPONSE						
Antigen presentation						
HLA-F	HLA class I histocompatibility antigen, alpha chain F	0.58	INF-γ	•		
HLA-E	HLA class I histocompatibility antigen, E alpha chain	0.65	INF-γ	•	•	
HLA-G	HLA class I histocompatibility antigen, alpha chain G	0.66	INF-γ/NF-κB	•		
Antimicrobial activity						
TFF1	Trefoil factor 1	0.52		•		
GBP1	Interferon-induced guanylate-binding protein 1	0.52	IFN-γ/LPS[f]	•	•	
LCN2	Lipocalin 2	0.57			•	
IGHA1	Ig alpha-1 chain C region	0.60		•		
SCGB3A1	Uteroglobin-related protein 2	0.69			•	
GBP5	Interferon-induced guanylate-binding protein 5	0.72	IFN-γ/LPS		•	
RNASEL	2-5A-dependent ribonuclease (EC 3.1.26.-)	0.78	INF-γ		•	
LEAP-2	Liver-expressed antimicrobial peptide 2	1.25			•	
DEFA6	Defensin 6	1.59			•	
PYY2	peptide YY, 2	1.60		•	•	
PRAP1	proline-rich acidic protein 1	1.87		•	•	
Cell adhesion						
SELE	Selectine E	0.05				•
CD36	Platelet glycoprotein IV	1.78		•	•	
SELP	Selectine P	2.44				•
Cell surface receptor						
NK4	Natural killer cells protein 4	0.61		•		
LST1/NKp30	Leukocyte specific transcript 1	0.68			•	
OCIL/LLT1	Osteoclast inhibitory lectin isoform 1	0.80			•	
KLRG1	Killer cell lectin-like receptor subfamily G, member 1	0.49			•	
TNFRSF11A	Tumor necrosis factor receptor superfamily member 11A	1.35			•	
Complement cascade						
C3	Complement component 3	0.40				•
DF	Complement factor D (EC 3.4.21.46)	1.44		•		
F10	Coagulation factor X precursor (EC 3.4.21.6)	1.98		•	•	
Cytokine and chemokine activity						
IL-1B	Interleukin-1B	0.01				•
IL-8	Interleukin-8	0.01				•
IL-17	Interleukin-17	0.03				•
LTA	Lymphotoxin alpha	0.04				•
CSF2	Colony stimulating factor 2	0.07				•
IL-6	Interleukin-6	0.08				•
CSF3	Colony stimulatory factor 3	0.14				•
IL-1A	Interleukin-1A	0.15				•
IL-15	Interleukin-15	5.26				•
Signal transduction						
JIP3	C-jun-amino-terminal kinase interacting protein 3	0.58		•	•	
MDS032	Putative MAPK activating protein PM26	0.68			•	
DUSP6/MKP	Mitogen-activated protein kinase phosphatase 3	0.68			•	
NFKBIZ	Nuclear factor kappa light polypeptide gene enhancer in B-cells inhibitor, zeta	0.71	TLR4[g]		•	
PPM1D	Protein phosphatase 2C delta isoform (EC 3.1.3.16)	0,79			•	
MAP3K7	Mitogen-activated protein kinase kinase kinase 7 (EC 2.7.1.-)	0.80	TLR4		•	

IRAK4	Interleukin-1 receptor-associated kinase 4	1.26	TLR			•	
Transcription factors							
CARM1	Coactivator-associated arginine methyltransferase 1	0.71				•	
ETS2	v-ets erythroblastosis virus E26 oncogene homolog 2	0.72				•	
EVI1	Ecotropic virus integration 1 site protein	0.78				•	
Miscellaneous							
NOS2A	Nitric oxide synthasa activity	0.09					•
NOD27	Nucleotide-binding oligomerization domains 27	0.51		•		•	
GZMB	Granzyme B (EC 3.4.21.79)	0.59		•		•	•
FLJ20241	putative NFkB activating protein	1.44				•	
CYP7A1	Cytochrome P450 family 7, subfamily A, polypeptide 1	3.13					•
PTGS2	Cyclooxygenase 2b	16.67					•
REGULATORY RESPONSE							
Cell growth							
Smad 7	SMAD, mothers against DPP homolog 7	2.94					•
Smad 3	SMAD, mothers against DPP homolog 3	6.67					•
Cell surface receptor							
CD19	CD19 antigen	0.39					•
CTLA4	Cytotoxic T-lymphocyte-associated protein 4	0.45					•
Chemiokine and cytokine activity							
IL-10	Interleukin-10	0.03					•
TNFRSF18	Tumor necrosis factor receptor superfamily member 18	6.67					•

a The regulation of these genes is based on information from literature; b MAANOVA: Microarray ANOVA analysis; c SC: Microarray standard correlation analysis; d LDA: Taqman Low Density Arrays; e Gene with both innate and adaptive functions; f LPS: lipopolysaccharide; g TLR: Toll-like receptor.

not shown). This suggests that IL-12 p35 is not involved in the activation of innate immunity through the functional form of IL-12, although it may have another yet unknown function in CD pathogenesis.

Concurrent with remission, a reduction in expression was observed for genes controlling the NF-κB signal cascade such as interleukin-1A, (*IL-1A*), interleukin-1B (*IL-1B*), interleukin-6 (IL-6), mitogen-activated protein kinase kinase kinase 1 (*TAK1*), natural killer cells protein 4 (*NK4*), nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, zeta (*NFKBIZ*), and the coactivator-associated arginine methyltransferase 1 (*CARM1*).

Of particular interest were the signals pointing to the persistent recruitment and activation of neutrophils during inflammation as suggested by the diet-induced attenuation of several potent chemokines: granulocyte-macrophage colony stimulating factor (*CSF2*), interleukine-1 (*IL1A* and *IL1B*), interleukine-8 (*IL8*), and lymphotoxin alpha (*LTA*), and the integrins selectin E (*SELE*), and selectin P (*SELP*) (Table 1). Leukotriene B4 (*LB4*) is a potent and specific activator of neutrophils and is degraded by the catabolic enzymes cytochrome P450, family 4, subfamily F, polypeptides 2 (CYP4F2) and 3 (CYP4F3). The expression of *CYP4F2* and *CYP4F3* was higher in both MIII and M0 patients than in controls (Fig. 3 and supplementary Fig. 2). This suggests that the necessity to negatively regulate LB4, and thus control neutrophil activity, is maintained despite the mucosal restoration on a gluten-free diet. Enhanced neutrophil infiltration in the lamina propria of coeliac patients was confirmed by immunohistochemistry using neutrophil granule-specific myeloperoxidase antibodies (Fig. 2B). Both MIII atrophic patients as well as M0 remission patients showed a two-fold increase of neutrophils compared to controls (Fig. 4 and supplementary Table 2). This suggests that despite the clinical, serological, and histological normalization, M0 patients showed an activated innate immunity that may reflect their genetic susceptibility towards relapsing with the disease.

The adaptive immune response

The key cells of the Th1 immune response in CD are the repertoire of gluten-restricted αβ αβ $CD4^+$ T cells in the lamina propria. Although the mechanism of T cell activation through HLA-DQ2

and -DQ8 gluten presentation is well understood, less is known about the molecular pathways that sustain and control this adaptive response.

Our observations of activation of the principal Th1 mediator IFNG, and its regulator T-box transcription factor TBX21 (*T-bet*) in the inflamed mucosa, were in agreement with earlier reports (Table 1).[8, 18] Activation downstream from IFNG followed a specific route in the pro-inflammatory JAK-STAT pathway involving signal transducer and activator of transcription 1-alpha/beta (*STAT-1*), tyrosine-protein kinase JAK2 (*JAK2*), and interferon consensus sequence binding protein (*IRF-8*). In total, we identified 19 differentially expressed genes downstream from IFNG (Table 1).

The chronic inflammatory response in CD is maintained by chemokines and cytokines that attract and induce activation, differentiation and proliferation of a variety of immune cell types. Increased expression of chemokine (C-X-C motif) ligand 10 (*CXCL10*) and –11 (*CXCL11*), chemokine (C-C motif) ligand 19 (CCL19) and chemokine (C-C motif) receptor-like 1 (*CCRL1*) may indicate recruitment and activation of T-cells, B-cells and γδ-T cells into the lamina propria during the active phase of the disease. In addition, the pro-inflammatory cytokines, interleukin-2 (*IL2*), its receptor interleukin-2 receptor A (*IL2RA*), interleukin-18 (*IL18*), the small inducible cytokines A4 (*CCL4*), A5 (*CCL5*) and A18 (*CCL18*) also increase during inflammation (Table 1).

Differential expression of genes involved in the regulation and suppression of the Th1 response pointed to the use of preferential receptor-ligand interactions in T cells and antigen-presenting cells (APCs). The APC ligand *CD86* was differentially expressed, together with its cognate T cell receptors cytotoxic T-lymphocyte-associated protein 4 (*CTLA4*), and inducible T-cell co-stimulator (*ICOS*). In contrast, no changes were measured in transcription for the APC ligand *CD80* and the T cell receptor *CD28*. The reduced expression of *IL10* during remission further underscored the Th1 suppressive activity of this regulatory cytokine during inflammation.[19]

An indication of involvement of the Th2 response was limited to decreased expression of the adaptor protein with pleckstrin homology and src homology 2 domains (*APS*), which is part of the B-cell receptor machinery, and an increased

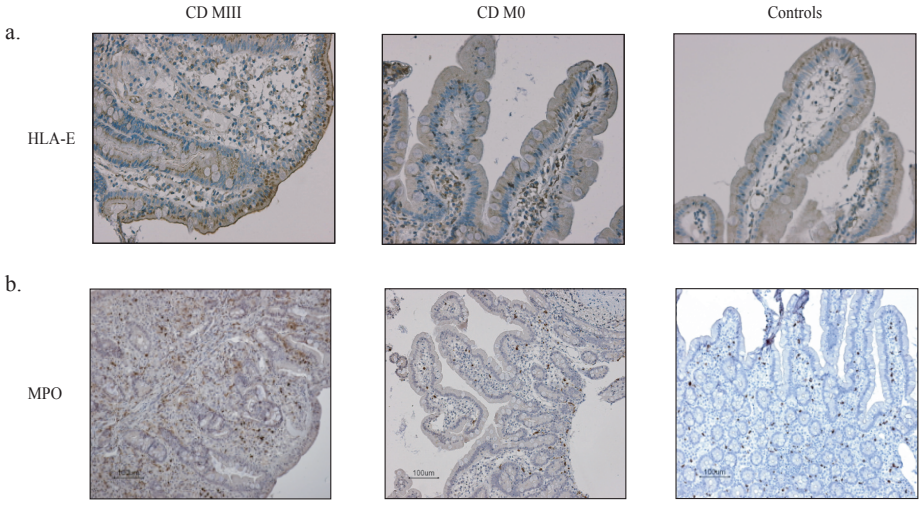


Figure 2. Immunohistochemical localization of HLA-E and myeloperoxidase (MPO) in the duodenal mucosa of coeliac patients and controls. Shown is the expression of the non-classical MHC class I molecule HLA-E and the neutrophil-specific marker MPO in patients (Marsh IIII and Marsh 0) and healthy controls. A) HLA-E was strongly expressed in the Marsh III epithelial layer but absent in normalized patients (Marsh 0) and controls. No differences in staining were found at the lamina propria. B) Increased number of MPO-positive cells in the lamina propria of Marsh III and Marsh 0 patients compared to controls. This indicated a higher number of neutrophils in the coeliac mucosa regardless of the histological stage or the gluten-free diet followed.

expression of antigen *CD19* that attenuates the signal through the B-cell receptor (Table 1). However, although important Th2 interleukin genes, such as *IL4* or *IL5*, were included in our experiments, they were below detection levels (results not shown). This suggests that, at the transcriptional level, there is no activation of the Th2 response in the coeliac mucosa. This does not preclude a response being evoked elsewhere as judged by the circulating antibodies directed against tissue transglutaminase and gliadin, for example.[20]

DISCUSSION

We present a genomics approach to the transcriptional regulation of immune-related genes in the intestinal mucosa of CD patients. The advantages of such an approach are that it is hypothesis-free and it provides a global assessment of gene regulation in the intricate molecular network of the immune system. Despite the fact that alternative ways of regulation will go unnoted (e.g. the release of compartmentalized proteins), determining transcriptional dynamics has the advantage that it can be accurately quantified with high-throughput technology (as opposed to immunohistochemistry, for example). We decided to study this process in whole mucosal biopsy samples from disease-affected tissues since these would reflect most accurately the *in vivo* situation. Intrinsic to our choice was that the observed expression profiles could result from the activation and recruitment of various cell types, which could also obscure minor changes within a less well represented cell type. This could make interpretation of the expression profiles precarious, but we supported our interpretation with immunohistochemistry where required. Nevertheless, the validation for this approach comes from the confirmation and further specification of coeliac-related immune processes reported earlier: the central role of the Th1 response;[3] the pivotal position of the pro-inflammatory cytokine IFNG (33 downstream genes detected);[18, 21] a lack of Th2 transcriptional activation;[22] and the involvement of innate immunity.[16, 17] Most expression studies are set up as case/control experiments and suffer from noise generated by individual differences in genetic background. This unavoidably leads to loss of information on more subtle molecular changes and will yield only a tip-of-the-iceberg view. The use of well characterized biopsy samples from patients in various stages of remission provided virtual time points along a recovery path that is more tolerant to individual variability. This allowed us to apply additional data analysis using Standard Correlation, which helped to select genes that showed subtle, but consistent, expression profiles that fell below the significance threshold of the MAANOVA analysis but which may still be physiologically relevant. We believe that the development of statistical methods that take temporal dependence between expression profiles into account is important to provide additional power to detect these small changes

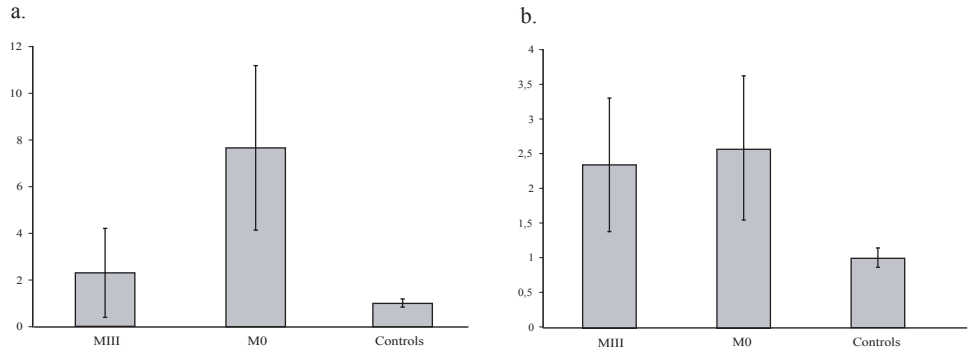


Figure 3. Relative expression of *CYP4F2* (A) and *CYP4F3* (B) determined by qRT-PCR. Gene expression was measured with qRT-PCR on RNA from individual M0 (n=16) and MIII (n=13) biopsy samples. Data were averaged and normalized to that obtained from a pool of cDNAs (n=16) derived from normal controls. Error bars indicate standard deviations. Measurements were in duplicate.

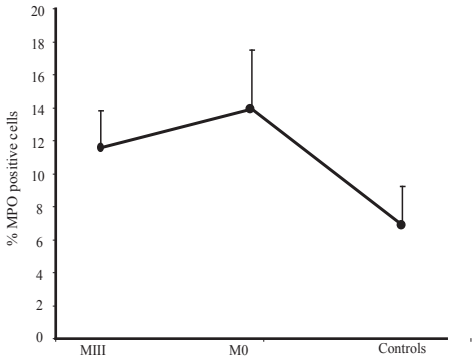


Figure 4. Graphical representation of the fraction of MPO-positive neutrophils counted in mucosal sections of coeliac patients (MIII and M0) and controls. On average a two-fold increase in neutrophils was observed in both active and remission patients compared to healthy controls. Percentages are based on the average of approximately 1000 total cells counted (see also supplementary table 2). Error bars indicate the standard error.

Our study has shown that during remission all immune-related genes, both innate and adaptive, as well as non-immune genes, increase or decrease their transcript levels in a gradual fashion. No significant gene expression profiles specific for one of the Marsh stages could be identified. This is in striking contrast with the histology of the mucosa that shows discrete transitions involving the sequential recovery from villous atrophy, crypt hyperplasia, and lymphocytosis. If the underlying molecular changes that drive this tissue restructuring are all gradual processes, it implies that the Marsh classification points to quantitative rather than qualitative differences between patients. This will have major consequences for diagnosing coeliac patients based on the presence or absence of MIII features. In this respect, individuals presenting with merely MII or even MI should also be considered coeliac patients as long as they demonstrate a positive response to a gluten-free diet. This substantiates a view based on the effects of a gluten-free diet on MII and MI patients.[23] It is therefore tempting to speculate that it is the combination of the period and amount of gluten exposure, on the one hand, and genetic susceptibility factors, on the other, that determine the amplitude of the immune response. This in turn will dictate whether the gluten-induced mucosal restructuring will extend all the way to the MIII stage or whether it stalls at MII or MI. Nonetheless, the atrophic MIII condition is the major cause of the chronic malabsorption-related pathology seen in coeliac patients.

Judged by the number of activated genes, and the various cell types and pathways involved, there appeared to be a comparable contribution of both the innate and adaptive immune response to the inflamed coeliac mucosa. This dual response might reflect the property of gluten to harbor both cytotoxic and immunogenic epitopes.[24] However, there is mounting evidence for the concerted and intertwined action of the innate and adaptive system in inflammatory diseases in general.[25]

In relation to innate immunity, there were several indicators for the chronic recruitment and activation of neutrophils: 1) enhanced cytokine transcript levels of *IL1*, *IL8*, *LTA*, and *CSF2*; 2) up-regulation of the selectins *SELE* and *SELP* on IL-1-stimulated endothelial cells; 3) activated leukotriene B4-degrading genes *CYP4F2* and *CYP4F3*; and 4) a two-fold increase of MPO-positive neutrophils. Neutrophils are short-lived phagocytes that circulate in large numbers in the bloodstream and are recruited to infected foci in the initial phase of inflammation. Although neutrophils activate anti-apoptotic processes upon stimulation, it was surprising to observe their mucosal involvement in a chronic, non-infectious disorder like CD. It should be noted that all patients refrain from eating for 12 hours before the biopsy is taken so that recent gluten exposure is highly unlikely. We found it even more remarkable that, judged by *CYP4F2/CYP4F3* expression and MPO-immunohistochemistry, neutrophils were involved in completely normalized M0 tissue of patients after an extensive period on a gluten-free diet. This all points to a coeliac-specific, but gluten- and Marsh lesion-independent, neutrophil infiltration and activation. There is a growing awareness that neutrophils may play an important role in the pathogenesis of inflammatory conditions: e.g. IBD[26] and asthma.[27] Neutrophil infiltration is generally associated with tissue damage due to the release of oxidative enzymes. Recent studies on neutrophil biology however dismiss this view since the granule release and physiological activity of these oxidative enzymes is limited to the intracellular vacuoles containing opsonized micro-organisms.[28] However, neutrophils reportedly have an influence on the

epithelial barrier by increasing its permeability.[29] The coeliac intestine is also characterized by an enhanced permeability, both in active as well as in remission patients.[30]

We recently reported identification of the *MYO9B* gene from the CELIAC4 locus that may confer genetic susceptibility by undermining the intestinal epithelial barrier.[6] Such an impaired barrier may provide access to the lamina propria of not just gluten but also of various other antigens and commensals. This may simultaneously facilitate gluten-presentation to αβ $CD4^+$ T cells and phagocytosis of microbes by neutrophils. Since neutrophil activation and the Th1 response both have a negative effect on epithelial barrier integrity, the intestinal permeability will be further increased. This results in a self re-enforcing mechanism of barrier impairment and a concerted innate/adaptive immune activation. Although this vicious circle can be interrupted by a gluten-free diet to attenuate the Th1 response, it will not be able to restore the genetic impairment of the barrier. A constitutive leaky intestinal barrier may not only contribute to CD, but also to various other inflammatory and autoimmune conditions. Recognition of impaired permeability as a common etiology of inflammatory disorders may help to identify additional causative genes and pave the way for therapeutic or prophylactic intervention.

APPENDIX

Supplementary data associated with this article can be found at http://humgen.med.uu.nl/publications/CD/Diosdado2006_1.html

ACKNOWLEDGEMENTS

We would like to thank Jackie Senior for editing the manuscript. We thank Joop van Helvoort and Philip Lijnzaad (Genomics Laboratory, UMC Utrecht) for technical support with microarrays and bioinformatics assistance, respectively. We are grateful to our former colleague Jan Geertzema for establishing the immunohistochemistry for our section. We are indebted to Jos Meijer (Pathology, Rijnstate Hospital, Arnhem) for biopsy sections and Marsh classification, Johan Offerhaus (Pathology, AMC, Amsterdam) for critical evaluation of immunohistochemistry results, and Maria Luisa Mearin (Pediatrics, LUMC, Leiden) for providing additional pediatric patient samples.

REFERENCES

1. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity (‘celiac sprue’). Gastroenterology 1992;102:330-54.

2. Diosdado B, Wapenaar MC, Franke L, Duran KJ, Goerres MJ, Hadithi M, Crusius JB, Meijer JW, Duggan DJ, Mulder CJ, Holstege FC Wijmenga C. A microarray screen for novel candidate genes in CD pathogenesis. Gut 2004;53:944-51.

3. Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2002;2:647-55.

4. van Heel DA, Hunt K, Greco L Wijmenga C. Genetics in coeliac disease. Best Pract Res Clin Gastroenterol 2005;19:323-39.

5. Londei M, Ciacci C, Ricciardelli I, Vacca L, Quaratino S Maiuri L. Gliadin as a stimulator of innate responses in celiac disease. Mol Immunol 2005;42:913-8.

6. 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 2005;37:1341-4.

7. 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.

8. Wapenaar MC, van Belzen MJ, Fransen JH, Sarasqueta AF, Houwen RH, Meijer JW, Mulder CJ Wijmenga C. The interferon gamma gene in celiac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. J Autoimmun 2004;23:183-90.

9. Roepman P, Wessels LF, Kettelarij N, Kemmeren P, Miles AJ, Lijnzaad P, Tilanus MG, Koole R, Hordijk GJ, van der Vliet PC, Reinders MJ, Slootweg PJ Holstege FC. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. Nat Genet 2005;37:182-6.

10. Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, Coulson R, Farne A, Lara GG, Holloway E, Kapushesky M, Lilja P, Mukherjee G, Oezcimen A, Rayner T, Rocca-Serra P, Sharma A, Sansone S Brazma A. ArrayExpress--a public repository for microarray gene expression data at

the EBI. Nucleic Acids Res 2005;33:D553-5.

11. van Bakel H, Strengman E, Wijmenga C Holstege FC. Gene expression profiling and phenotype analyses of S. cerevisiae in response to changing copper reveals six genes with new roles in copper and iron metabolism. Physiol Genomics 2005;22:356-67.

12. Franke L, van Bakel H, Diosdado B, van Belzen M, Wapenaar M Wijmenga C. TEAM: a tool for the integration of expression, and linkage and association maps. Eur J Hum Genet 2004;12:633-8.

13. Livak KJ Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001;25:402-8.

14. Wu H KK, Xiangqin C, and Churchill GA. MAANOVA: A software package for the analysis of spotted cDNA microarray experiments. In: Parmigiani G, Garrett, E. S., Irizarry, R. A., and Zeger, S. L., ed. The analysis of gene expression data: Methods and Software. Springer-Verlag, New York, 2003.

15. Borrego F, Masilamani M, Kabat J, Sanni TB Coligan JE. The cell biology of the human natural killer cell CD94/NKG2A inhibitory receptor. Mol Immunol 2005;42:485-8.

16. Hue S, Mention JJ, Monteiro RC, Zhang S, Cellier C, Schmitz J, Verkarre V, Fodil N, Bahram S, Cerf-Bensussan N Caillat-Zucman S. A direct role for NKG2D/MICA interaction in villous atrophy during celiac disease. Immunity 2004;21:367-77.

17. Meresse B, Chen Z, Ciszewski C, Tretiakova M, Bhagat G, Krausz TN, Raulet DH, Lanier LL, Groh V, Spies T, Ebert EC, Green PH Jabri B. 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.

18. Monteleone I, Monteleone G, Del Vecchio Blanco G, Vavassori P, Cucchiara S, MacDonald TT Pallone F. Regulation of the T helper cell type 1 transcription factor T-bet in coeliac disease mucosa. Gut 2004;53:1090-5.

19. Salvati VM, Mazzarella G, Gianfrani C, Levings MK, Stefanile R, De Giulio B, Iaquinto G, Giardullo N, Auricchio S, Roncarolo MG Troncone R. Recombinant human interleukin 10 suppresses gliadin dependent T cell activation in ex vivo cultured coeliac intestinal mucosa. Gut 2005;54:46-53.

20. Rostami K, Mulder CJ, Stapel S, von Blomberg BM, Kerckhaert J, Meijer JW, Pena SA Heymans HS. Autoantibodies and histogenesis of celiac disease. Rom J Gastroenterol 2003;12:101-6.

21. Nilsen EM, Jahnsen FL, Lundin KE, Johansen FE, Fausa O, Sollid LM, Jahnsen J, Scott H Brandtzaeg P. Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease. Gastroenterology 1998;115:551-63.

22. Forsberg G, Hernell O, Melgar S, Israelsson A, Hammarstrom S Hammarstrom ML. Paradoxical coexpression of proinflammatory and down-regulatory cytokines in intestinal T cells in childhood celiac disease. Gastroenterology 2002;123:667-78.

23. PJ Wahab Coeliac Disease, more than villous atrophy. Arnhem, the Netherlands, 2002.

24. Ciccocioppo R, Di Sabatino A Corazza GR. The immune recognition of gluten in coeliac disease. Clin Exp Immunol 2005;140:408-16.

25. Vivier E Malissen B. Innate and adaptive immunity: specificities and signaling hierarchies revisited. Nat Immunol 2005;6:17-21.

26. Kucharzik T, Walsh SV, Chen J, Parkos CA Nusrat A. Neutrophil transmigration in inflammatory bowel disease is associated with differential expression of epithelial intercellular junction proteins. Am J Pathol 2001;159:2001-9.

27. Kamath AV, Pavord ID, Ruparelia PR Chilvers ER. Is the neutrophil the key effector cell in severe asthma? Thorax 2005;60:529-30.

28. Segal AW. How neutrophils kill microbes. Annu Rev Immunol 2005;23:197-223.

29. Edens HA, Levi BP, Jaye DL, Walsh S, Reaves TA, Turner JR, Nusrat A Parkos CA. Neutrophil transepithelial migration: evidence for sequential, contact-dependent signaling events and enhanced paracellular permeability independent of transjunctional migration. J Immunol 2002;169:476-86.

30. 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;43:435-41.

Part III

CANDIDATE GENES STUDIES

Chapter 5

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

Am J Physiol Gastrointest Liver Physiol.
2005 Sept;289(3):G495-500

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.

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).

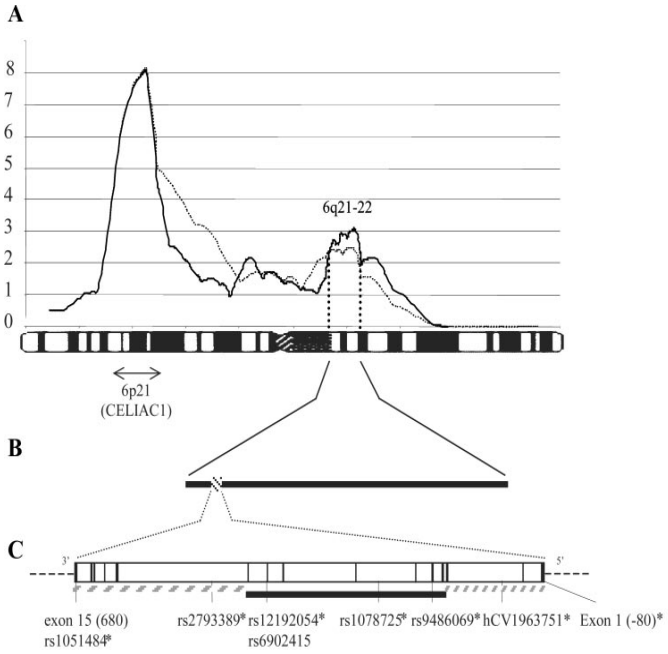
Sequence analysis. To investigate whether the enzymatic properties of this gene product or its expression levels were different in CD patients due to an underlying genetic variation, we performed sequence analysis on the entire coding region and exon-intron boundaries of the *PREP* gene to identify putative 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).

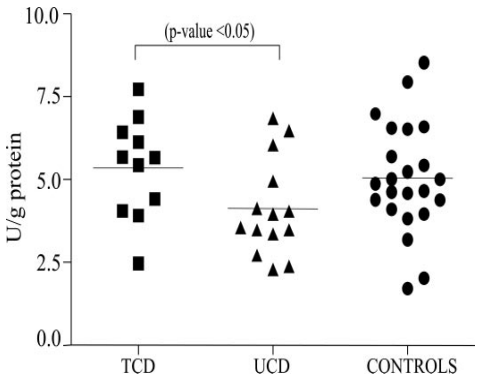


Fig. 2. PREP activity in duodenal biopsies. The activity was measured with Z-Gly-Pro-AMC substrate and corrected for protein concentration determined with BCA assay. TCD 5.35 ± 0.46; UCD 4.12 ± 0.39; $P < 0.05$; controls 5.03 ± 0.033. TCD, treated celiac disease (gluten-free diet); UCD, untreated celiac 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.

Chapter 6

The down-stream modulator of IFNG, STAT1 is not genetically associated to the Dutch Coeliac disease population

Submitted

The down-stream modulator of interferon-gamma, STAT1 is not genetically associated to the Dutch coeliac disease population

Begoña Diosdado[1*], Alienke J Monsuur[1*], Maria Luisa Mearin[2,3], Chris Mulder[4] and Cisca Wijmenga[1]

[1] Complex Genetics Section, Department of Biomedical Genetics, University Medical Centre, Utrecht, [2] Pediatric Gastroenterology, Leiden University Medical Center, Leiden, [3] VU University Medical Centre, Amsterdam, [4]Department of Gastroenterology, VU University Medical Centre, Amsterdam, the Netherlands

** These authors contributed equally to this work.*

Coeliac disease (CD) is a complex genetic disorder. Its etiology is due to multiple genes and environmental factors, such as gluten. The first event in the pathogenesis of CD after the ingestion of gluten is the activation of a Th1 immune response that leads to villous atrophy. Although this immune response seems crucial to the disease's development, only the HLA-*DQ2/DQ8* genes have been identified as causative immune genes related to CD. Recently, the activation of the transcription factor STAT1 and changes in its expression levels have confirmed the participation of the JAK-STAT pathway in CD. Furthermore, since the *STAT-1* gene is a positional candidate located in the CELIAC3 locus on chromosome 2, we speculate that alterations in this gene could be primarily responsible for the aberrant immune response that characterizes CD. Based on this functional and genetic evidence, we investigated the primary contribution of *STAT-1* to CD. We performed a comprehensive genetic association study using five tag SNPs fully covering the *STAT-1* gene in a Dutch cohort of 355 independent CD cases and 360 healthy controls. Neither the alleles, nor the genotypes in the case-control genetic association studies, nor the haplotype analysis showed any association to the *STAT-1* gene in the Dutch CD population. Our results do not point to a primary involvement of the *STAT-1* gene in the Dutch CD population.

INTRODUCTION

Coeliac disease (CD) is a complex genetic disorder in which the main causative environmental factor is the dietary protein gluten. This is present in common Western foods, such as bread, pasta and many processed products. When gluten is ingested by coeliac patients, it provokes an inflammatory reaction in the small intestine that leads to crypt hyperplasia and villous atrophy [1]. Although the exact nature of this immune response remains unknown, there is much evidence to suggest that gluten promotes a Th1 immune response mostly mediated by high levels of interferon-gamma (IFN-γ) [2-13] and αβ CD4+ T cells that recognize gluten peptides through HLA-DQ2/DQ8 molecules [14]. This observation is further supported by genetic data showing a primary association between the genes encoding the HLA-DQ2/DQ8 molecules on chromosome 6 and all the CD populations tested to date [15, 16]. Despite the fact that these results indicate an essential role for HLA in the etiopathogenesis of CD, the HLA-DQ2/DQ8 genes only explain approximately 40% of the familial clustering to CD [16, 17], implying that other non-HLA genes must also be associated with CD. Since the Th1 immune response is such a crucial factor in the pathogenesis of CD, the search for other primary genes should focus on genes linked to the Th1 immune response.

The Janus kinase-signal transducer and activator of transcription (JAK-STAT) pathway plays an important role in regulating the immune response. It mediates the activity of several cytokines by

activating gene transcription through different STAT proteins [18]. We can hypothesize that, in CD, a disturbed activity of the JAK-STAT pathway might lead to an altered transcription of cytokines in the intestine of coeliac patients. One interesting activator of the JAK-STAT pathway in CD might be IFNG, which is also highly expressed in the intestine of coeliac individuals [2-9]. The exact role of IFNG in the pathogenesis of CD has not been elucidated, but it might contribute to the maintenance of the chronic inflammatory response by activating the JAK-STAT pathway through STAT1.

Although IFNG might play an important role in CD pathogenesis [2-9], genetic association studies have failed to show that this gene makes a primary contribution to the disease [2, 19]. Interestingly, its direct down-stream target, the STAT1 protein has been shown to be activated in the mononuclear cells of the lamina propria and epithelial cells of CD biopsies [20]. Furthermore, the *STAT1* gene is both a positional and functional candidate gene, since it is located in the CELIAC3 locus [21] and mRNA expression studies on biopsies of CD patients have shown that *STAT1* is expressed more highly in the intestinal mucosa of patients than in controls (Diosdado et al. in preparation). These observations suggest that STAT1 might be critical in controlling the immunological activity in CD and, hence, any genetic change that alters its transcriptional activity could explain the detrimental immune response seen in CD. Based on this evidence, we carried out a comprehensive approach using tag SNPs to investigate the genetic contribution of the non-HLA gene *STAT1* in the Dutch CD population in a case-control cohort.

METHODS

DNA samples for genetic association studies

For the genetic association study we collected a group of 355 independent CD cases and 360 independent blood bank controls, all of Dutch Caucasian origin. All CD patients were diagnosed according to the ESPGHAN criteria [22] and were selected for showing a Marsh III lesion at initial diagnosis. Ninety-three percent of the patients were HLA-DQ2 positive. We collected blood samples and isolated DNA according to standard laboratory procedures [16].

Informed consent was obtained from all individuals and the study was approved by the Medical Ethics Committee of the University Medical Centre, Utrecht, the Netherlands.

SNP selection and data analysis

To select the tag SNPs we downloaded the genotype information on the SNPs for the *STAT1* gene from the HapMap database (http://www.hapmap.org/) (typed in the CEPH population of Utah residents with ancestors from northern and western Europe). We uploaded this to the Tagger program (http:www.broad.mit.edu/mpg/tagger/). This software returned six non-redundant tag SNPs with a minor allele frequency (MAF) $\geq$ 10% and an $r^2 \geq 0.7$ that covered the *STAT1* gene. For these SNPs, two assay-on-demand probes: rs3771300 (C__22274194_10) and rs1914408 (C_1737196_10), and four assay-by-design probes: rs7562024 (C__1737208_10), rs1547550 (C__1737199_10), rs13010343 (C__1737197_10) and rs2030171 (C__1737225_20) were obtained. All the tested SNPs were analyzed on an ABI Prism 7900

Table 1. P-values obtained from the genetic association study in the *STAT1* gene using tag SNPs on 355 CD patients and 360 controls.

a.

SNP rs number	Bp position	SNP type	MAF*	Cases (%)	Controls (%)	Chi-square	p-value
rs3771300	191661102	T/G	G	48.0	48.5	1.78	0.1821
rs1914408	191665482	C/T	T	25.5	24.9	0.078	0.7807
rs13010343	191668951	A/G	A	15.8	12.4	3.36	0.0668
rs7562024	191681027	C/T	T	35.9	37.1	0.226	0.6344
rs2030171	191694669	A/G	A	32.9	31.8	0.206	0.6502

b.

SNP rs number	Genotype	Cases (%)	Controls (%)	Chi-square*	p-value**
rs3771300	GG	89 (25.6)	96 (26.8)		
	GT	155 (44.7)	177 (49.4)		
	TT	103 (29.7)	85 (23.7)		
				1.68	0.19
rs1914408	CC	194 (55.9)	200 (56.8)		
	CT	129 (37.2)	129 (36.6)		
	TT	24 (6.9)	23 (6.5)		
				0.00	0.96
rs13010343	AA	250 (71.8)	271 (77.2)		
	AG	86 (24.7)	73 (20.8)		
	GG	12 (3.4)	7 (2.0)		
				1.64	0.20
rs7562024	CC	145 (41.8)	142 (40.7)		
	CT	155 (44.7)	155 (44.4)		
	TT	47 (13.5)	52 (14.9)		
				0.03	0.87
rs2030171	AA	35 (10.2)	38 (10.9)		
	AG	155 (45.3)	145 (41.7)		
	GG	152 (44.4)	165(47.4)		
				0.24	0.63

* minor allele frequency, **overall chi-square p-value of all three genotypes.(a)Tag SNP information and association studies on allele frequencies. (b) Association studies on genotype frequencies.

HT system. All reagents and equipment were purchased from Applied Biosystems (Foster, CA, USA) and used according to their protocols.

Hardy-Weinberg equilibrium was evaluated separately in the case and control groups, for all tested SNPs (data not shown). Differences in allele frequencies and genotypes were compared between cases and controls using a χ^2 test.

RESULTS

To study the *STAT1* gene as a causative factor in CD, we selected six tag SNPs for further genetic association studies (Table 1). Although the SNP rs1547550 could not be tested due to manufacturing failures, the other five tag SNPs captured 90% of the untyped SNPs present in the gene with an $r^2 > 0.7$ and MAF > 10%. All SNPs were in HWE (data not shown). The association studies showed no statistical difference between the cases and controls for either allele (Table 1a) or genotype frequencies (Table 1b). Haplotype analysis did not improve the results (data not shown). We therefore concluded that the *STAT1* gene is not associated to CD in the Dutch population.

DISCUSSION

The STAT1 protein is one of the main down-stream modulators of IFNG signaling through the JAK-STAT pathway. The *STAT1* gene is interesting in CD because it is located in the CELIAC3 locus on chromosome 2q23-32. This region has been found to be significantly linked to CD in the Scandinavian population and was also reported in a meta-analysis carried out by a European consortium on CD [15, 21]. This locus contains a number of interesting candidate genes for CD, but so far attention has mainly been focused on the *CTLA-4* gene, since it has shown association to other autoimmune disorders [23]. However, genetic studies in CD have been rather inconclusive and do not clearly point to a genetic association with *CTLA-4* [21, 24-37]. It is tempting to speculate that another gene in this region may be the causative one in CD pathogenesis.

We therefore conducted genetic association studies in a Dutch case-control cohort to determine the contribution of the *STAT1* gene in

the etiology of CD. Our association studies did not reach statistical significance for any of the tested tag SNPs. Our results gave no evidence for *STAT1* being a causative factor in the Dutch CD population.

Despite the results from the genetic studies on *IFNG* and *STAT1*, the importance of the JAK-STAT1 pathway in the immune response to gluten is still remarkable. Unfortunately, there have been no functional experiments to determine the exact nature of the stimuli that initiate the JAK-STAT1 pathway in CD. It would be logical to assume that the major stimulator of this pathway is IFNG, since large amounts of it are seen in the biopsies of CD patients, but other cytokines such as IL-6 can also activate this pathway [38]. Additional studies on the JAK-STAT1 pathway, revealing the range of genes regulated by this transcription factor under the presence of gluten, are important, and should reveal new aspects of the immune response that occurs in CD. We still need to determine whether other STAT (STAT1-6) and Janus kinase (JAK) proteins are activated in the coeliac intestine, and especially, which molecules of the regulatory mechanisms of this pathway (the suppressor of cytokine signaling (SCOCS) family, the protein inhibitor of activated STAT (PIAS) proteins and protein tyrosine phosphatases (PTPs)) participate in the pathogenesis of CD. Interestingly, some of the molecules of the regulatory mechanisms have already been related to immune diseases, like the regulatory molecule protein, tyrosine phosphatase receptor type, C (PTPRC), that has been associated to multiple sclerosis [39]. Identifying other members of the JAK-STAT pathway participating in CD, and the molecules that maintain its balance, might therefore reveal potential candidate genes for further genetic association studies in CD.

In conclusion, although functional studies imply that STAT1 plays an important role in regulating the extensive activation of the adaptive immune response in the lamina propria of CD, genetic association studies do not show any significant association of this gene to the Dutch CD population.

ACKNOWLEDGEMENTS

We would like to thank Jackie Senior for editing the manuscript and the lab members of the coeliac disease group for fruitful discussions. We thank the Netherlands Organization for Medical and Health Research (ZonMW, grants 902-22-094 and 912-02-028), and the CDC (Coeliac Disease Consortium), an innovative cluster approved by the Netherlands Genomics Initiative and partially funded by a Dutch government grant (BSIK03009), for financial support.

REFERENCES

1. Marsh MN. Gluten, major histocompatibility co mplex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity (‘celiac sprue’). Gastroenterology 1992;102:330-54.

2. Wapenaar MC, van Belzen MJ, Fransen JH, Sarasqueta AF, Houwen RH, Meijer JW, Mulder CJ, Wijmenga C. The interferon gamma gene in celiac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. J Autoimmun 2004;23:183-90.

3. Forsberg G, Hernell O, Melgar S, Israelsson A, Hammarstrom S, Hammarstrom ML. Paradoxical coexpression of proinflammatory and down-regulatory cytokines in intestinal T cells in childhood celiac disease. Gastroenterology 2002;123:667-78.

4. al-Dawoud A, Nakshabendi I, Foulis A, Mowat AM. Immunohistochemical analysis of mucosal gamma-interferon production in coeliac disease. Gut 1992;33:1482-6.

5. Nilsen EM, Jahnsen FL, Lundin KE, Johansen FE, Fausa O, Sollid LM, Jahnsen J, Scott H, Brandtzaeg P. Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease. Gastroenterology 1998;115:551-63.

6.Perez-Machado MA, Ashwood P, Thomson M A, Latcham F, Sim R, Walker-Smith JA, Murch SH. Reduced transforming growth factor-beta1-producing T cells in the duodenal mucosa of children with food allergy. Eur J Immunol 2003;33:2307-15.

7. Westerholm-Ormio M, Garioch J, Ketola I, Savilahti E. Inflammatory cytokines in small intestinal mucosa of patients with potential coeliac disease. Clin Exp Immunol 2002;128:94-101.

8. Veres G, Westerholm-Ormio M, Kokkonen J, Arato A, Savilahti E. Cytokines and adhesion molecules in duodenal mucosa of children with delayed-type food allergy. J Pediatr Gastroenterol Nutr 2003;37:27-34.

9. Kontakou M, Sturgess RP, Przemioslo RT, Limb GA, Nelufer JM, Ciclitira PJ. Detection of interferon gamma mRNA in the mucosa of patients with coeliac disease by in situ hybridisation. Gut 1994;35:1037-41.

10. Breese EJ, Kumar P, Farthing MJ, MacDonald TT.

Interleukin-2 and interferon-gamma producing cells in the lamina propria in celiac disease. Dig Dis Sci 1994;39:2243.

11. Salvati VM, Mazzarella G, Gianfrani C, Levings MK, Stefanile R, De Giulio B, Iaquinto G, Giardullo N, Auricchio S, Roncarolo MG, Troncone R. Recombinant human interleukin 10 suppresses gliadin dependent T cell activation in ex vivo cultured coeliac intestinal mucosa. Gut 2005;54:46-53.

12. Monteleone I, Monteleone G, Del Vecchio Blanco G, Vavassori P, Cucchiara S, MacDonald TT, Pallone F. Regulation of the T helper cell type 1 transcription factor T-bet in coeliac disease mucosa. Gut 2004;53:1090-5.

13. Lahat N, Shapiro S, Karban A, Gerstein R, Kinarty A, Lerner A. Cytokine profile in coeliac disease. Scand J Immunol 1999;49:441-6.

14. Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2002;2:647-55.

15. 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 2003;11:828-34.

16. 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.

17. Bevan S, Popat S, Braegger CP, Busch A, O'Donoghue D, Falth-Magnusson K, Ferguson A, Godkin A, Hogberg L, Holmes G, Hosie KB, Howdle PD, Jenkins H, Jewell D, Johnston S, Kennedy NP, Kerr G, Kumar P, Logan RF, Love AH, Marsh M, Mulder CJ, Sjoberg K, Stenhammer L, Walker-Smith J, Marossy AM, Houlston RS. Contribution of the MHC region to the familial risk of coeliac disease. J Med Genet 1999;36:687-90.

18. Kisseleva T, Bhattacharya S, Braunstein J, Schindler CW. Signaling through the JAK/STAT pathway, recent advances and future challenges. Gene 2002;285:1-24.

19. Woolley N, Mustalahti K, Maki M, Partanen J. Cytokine gene polymorphisms and genetic association with coeliac disease in the Finnish population. Scand J Immunol 2005;61:51-6.

20. Mazzarella G, MacDonald TT, Salvati VM, Mulligan P, Pasquale L, Stefanile R, Lionetti P, Auricchio S, Pallone F, Troncone R, Monteleone G. Constitutive activation of the signal transducer and activator of transcription pathway in celiac disease lesions. Am J Pathol 2003;162:1845-55.

21. Rioux JD, Karinen H, Kocher K, McMahon SG, Karkkainen P, Janatuinen E, Heikkinen M, Julkunen R, Pihlajamaki J, Naukkarinen A, Kosma VM, Daly MJ, Lander ES, Laakso M. Genomewide search and association studies in a Finnish celiac disease population: Identification of a novel locus and replication of the HLA and CTLA4 loci. Am J Med Genet A 2004;130:345-50.

22. 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. Kristiansen OP, Larsen ZM, Pociot F. CTLA-4 in autoimmune diseases--a general susceptibility gene to autoimmunity? Genes Immun 2000;1:170-84.

24. Hunt KA, McGovern DP, Kumar PJ, Ghosh S, Travis SP, Walters JR, Jewell DP, Playford RJ, van Heel DA. A common CTLA4 haplotype associated with coeliac disease. Eur J Hum Genet 2005;13:440-4.

25. Naluai AT, Nilsson S, Samuelsson L, Gudjonsdottir AH, Ascher H, Ek J, Hallberg B, Kristiansson B, Martinsson T, Nerman O, Sollid LM, Wahlstrom J. The CTLA4/CD28 gene region on chromosome 2q33 confers susceptibility to celiac disease in a way possibly distinct from that of type 1 diabetes and other chronic inflammatory disorders. Tissue Antigens 2000;56:350-5.

26. 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.

27. Popat S, Hearle N, Hogberg L, Braegger CP, O'Donoghue D, Falth-Magnusson K, Holmes GK, Howdle PD, Jenkins H, Johnston S, Kennedy NP, Kumar PJ, Logan RF, Marsh MN, Mulder CJ, Torinsson Naluai A, Sjoberg K, Stenhammar L, Walters JR, Jewell DP, Houlston RS. Variation in the CTLA4/CD28 gene region confers an increased risk of coeliac disease. Ann Hum Genet 2002;66:125-37.

28. Clot F, Fulchignoni-Lataud MC, Renoux C, Percopo S, Bouguerra F, Babron MC, Djilali-Saiah I, Caillat-Zucman S, Clerget-Darpoux F, Greco L, Serre JL. Linkage and association study of the CTLA-4 region in coeliac disease for Italian and Tunisian populations. Tissue Antigens 1999;54:527-30.

29. Holopainen P, Arvas M, Sistonen P, Mustalahti K, Collin P, Maki M, Partanen J. CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to celiac disease. A linkage and family-based association study. Tissue Antigens 1999;53:470-5.

30. King AL, Moodie SJ, Fraser JS, Curtis D, Reid E, Dearlove AM, Ellis HJ, Ciclitira PJ. CTLA-4/CD28 gene region is associated with genetic susceptibility to coeliac disease in UK families. J Med Genet 2002;39:51-4.

31. King AL, Moodie SJ, Fraser JS, Curtis D, Reid E,

Dearlove AM, Ciclitira PJ. Coeliac disease: investigation of proposed causal variants in the CTLA4 gene region. Eur J Immunogenet 2003;30:427-32.

32. Mora B, Bonamico M, Indovina P, Megiorni F, Ferri M, Carbone MC, Cipolletta E, Mazzilli MC. CTLA-4 +49 A/G dimorphism in Italian patients with celiac disease. Hum Immunol 2003;64:297-301.

33. Popat S, Hearle N, Wixey J, Hogberg L, Bevan S, Lim W, Stenhammar L, Houlston RS. Analysis of the CTLA4 gene in Swedish coeliac disease patients. Scand J Gastroenterol 2002;37:28-31.

34. Martin-Pagola A, Perez de Nanclares G, Vitoria JC, Bilbao JR, Ortiz L, Zubillaga P, Castano L. No association of CTLA4 gene with celiac disease in the Basque population. J Pediatr Gastroenterol Nutr 2003;37:142-5.

35. 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.

36. Amundsen SS, Naluai AT, Ascher H, Ek J, Gudjonsdottir AH, Wahlstrom J, Lie BA, Sollid LM. Genetic analysis of the CD28/CTLA4/ICOS (CELIAC3) region in coeliac disease. Tissue Antigens 2004;64:593-9.

37. Haimila K, Smedberg T, Mustalahti K, Maki M, Partanen J, Holopainen P. Genetic association of coeliac disease susceptibility to polymorphisms in the ICOS gene on chromosome 2q33. Genes Immun 2004;5:85-92.

38. Shuai K, Liu B. Regulation of JAK-STAT signalling in the immune system. Nat Rev Immunol 2003;3:900-11.

39. Jacobsen M, Schweer D, Ziegler A, Gaber R, Schock S, Schwinzer R, Wonigeit K, Lindert RB, Kantarci O, Schaefer-Klein J, Schipper HI, Oertel WH, Heidenreich F, Weinshenker BG, Sommer N, Hemmer B. A point mutation in PTPRC is associated with the development of multiple clerosis. Nat Genet 2000;26:495-9.

Part IV

GENERAL DISCUSSION

Aetiopathogenesis of coeliac disease

At the time we started the research described in this thesis in 2001, coeliac disease (CD) was one of the "best-characterised" of the complex disorders. The contribution of the major histocompatibility complex (MHC) locus (HLA-DQ) to the aetiopathogenesis of the disease was understood, and gluten was known to be the main environmental factor triggering the disease. CD was generally described as a complex genetic disorder in which gluten intake evokes an immune response in the small intestine that triggers crypt hyperplasia and villous atrophy. These histological changes were used to classify the disease into three stages: a first stage, known as Marsh I (MI), characterised by lymphocytosis, a second stage, Marsh II (MII), shows crypt hyperplasia together with lymphocytosis, and a last stage, Marsh III (MIII), that shows villous atrophy in addition to the Marsh II characteristics [1]. Since the genetic and pathogenic contribution of the HLA-DQ2/DQ8 molecules was established [2], the main research focus in the CD field was on the detailed characterisation of the adaptive immune reaction to gluten. In 2001 it was thought that a T-helper 1 (Th1) immune response would be the predominant factor in the immunopathogenesis. The Th1 response is mediated by a repertoire of αβ CD4+ T cells that become activated upon recognition of immuno-dominant gluten peptides, deamidated by tissue transglutaminase, and presented by HLA-DQ2/DQ8 molecules in the lamina propria [3]. However, the *HLA-DQ2/DQ8* genes could only explain 40% of the familial aggregation, and high levels of a few cytokines, such as interferon-gamma (IFN-γ) or tumour necrosis factor-alpha (TNF-α) had been described. Hence, the functionality of this Th1 immune response in the pathogenesis of the disease was largely unknown, and any attempt to further understand other aspects of the molecular pathology of CD was by trying to identify genes outside the MHC region (Fig. 1a) [4-9]. To find such genes, several genome screens were conducted and identified significant and suggestive regions of linkage. However, these studies did not identify other primary genes, and only showed controversial results for the *CTLA-4* gene, as reviewed in chapter 1.

A second approach for identifying causative genes was to follow a functional candidate gene approach, but this did not yield any fruitful results either [5, 10-12]. As shown in chapter 1, it was clear that the genetic and functional studies had put too much emphasis on the Th1 immune response. It was therefore necessary to explore other aspects of the pathogenesis of the disease using functional studies, genome-wide technologies such as microarray studies, genome-wide genetic association studies, or bioinformatics tools to integrate raw genetic and expression data as described in chapter 2.

In CD, functional studies can be carried out on biopsies from the duodenal mucosa where the pathogenesis of the disease takes place, since the site of the lesion is easily accessible and small intestinal biopsies need to be taken as part of the standard diagnostic procedure. These types of investigations led to the identification of a role for the innate immune response in the pathogenesis of CD through a non-HLA-driven immune reaction to a toxic gluten peptide [13]. This observation was further supported by experiments that revealed that this toxic peptide stresses epithelial cells that then express MHC class I polypeptide-related sequence A (MICA) molecules and produce IL-15, which stimulates lymphocyte activated killer (LAK)

cells expressing NKG2D molecules [14, 15]. Consequently, these experiments pointed not to the existence of a receptor that would recognise gluten and activate the innate response, but rather to gluten inducing epithelial stress in an unspecified manner (chapter 1). In the same studies, it was also suggested that villous atrophy is mediated by the innate immune response through the recognition of MICA and NKG2D molecules. This disagreed with other studies that described the adaptive immune response as mediating villous atrophy through the FAS-FAS-ligand pathway, or through its influence on the expression of metalloproteinases (MMPs) and their inhibitors (TIMPs) in the lamina propria, as explained in more detail in chapter 2 (Fig. 1a) [16].

Given this controversy and the lack of detailed insight into the molecular mechanisms involved in the pathology of CD, we reasoned that a hypothesis-free technology such as genome-wide expression profiling using microarrays (see Box 1) could be a powerful tool to identify: i) novel pathways and cell types involved in the pathogenesis of CD, ii) markers for the disease progression that could be used in the clinical setting, and iii) putative primary disease genes by the integration of gene expression data with genetic data.

The results of our genome-wide expression studies are described in chapters 3 and 4. The expression study described in chapter 3 aimed to unravel the molecular mechanisms that drive crypt hyperplasia and villous atrophy. It was performed on duodenal biopsies of MIII CD patients and controls. The results did not show any evidence for the previously proposed mechanisms for driving villous atrophy (chapter 2), but nonetheless they led us to propose a novel mechanism responsible for tissue remodelling in CD. We hypothesised that

BOX 1

Gene expression microarrays are a powerful technique that allows the retrieval of expression levels of thousands of genes simultaneously. Microarrays are microscope glass surfaces or nylon membranes on which cDNA or oligonucleotides are printed. The principle of this technique is the hybridisation of fragments of the genes present on the target samples to their complementary spotted probes [34]. Handling the large amount of information produced by microarray experiments in a comprehensive manner is a challenging task, and its success depends on the robustness of the experiment design, sample collection, experimental procedures, and data analysis and interpretation techniques. These studies have two main applications: one in the clinical setting used to classify diseases, and a second one in the research field to define expression levels to gain insight into the molecular mechanisms and function of genes. The first option does not take into account the functional relevance of the differentially expressed genes. Many successful examples of this technique have been published, mainly in the field of cancer, such as classification of distinct types of lymphomas [35], establishing prognosis information [36], or therapeutic responses [37]. In addition to clinical applications, microarray studies also provide insights into disease mechanisms, into biological processes and transcriptional regulation by generating transcriptional profiles, and help determine gene functions [38]. However, microarray studies only identify genes that are regulated at a translational level and, hence, they do not detect post-transcriptional regulated genes. These diverse applications of microarray studies are a complementary and alternative tool for investigating complex genetic disorders.

villous atrophy and crypt hyperplasia could be explained by an impaired ratio between the proliferation and terminal differentiation of the enterocytes. In addition, to gain more details of the inflammatory response in CD, we conducted another expression study in which we compared the expression profiles of biopsy materials from coeliac patients in different Marsh stages of the disease and from controls (chapter 4). This allowed us to identify differentially expressed genes and their dynamics during this mucosal restructuring. These experiments showed that the consecutive Marsh stages seen at the histological level are not reflected in the molecular process. The disease process seems to progress more gradually (chapter 4, Fig. 1). These findings also revealed important aspects of the adaptive response and the activation of neutrophils in the coeliac lesion even in completely normalized mucosa (M0) due to a gluten-free diet. This observation points to a gluten-independent neutrophil activation that might reflect the enhanced permeability of the intestinal barrier in CD patients due to a genetic predisposition as it will be discuss later [17]. Finally, the integration of the gene expression data and genetic linkage data highlighted two candidate genes for genetic association studies: prolyl-endopeptidase (*PREP*) and the signal transducer and activator of transcription 1-alpha/beta (*STAT1*) (chapters 5 and 6). These new pieces of the puzzle and their implications will be discussed in more detail later.

Novel pathways and cell types involved in the pathogenesis of CD

New insights into the immunopathogenesis of coeliac disease

The innate immune response is the first line of defence of the human body, and it is located in the epithelial layer of the small intestine. In CD, the innate response is activated through IL-15, which activates intraepithelial lymphocytes that express the receptor CD94/NKG2D and recognise MICA molecules expressed on epithelial cells [14, 15]. These *in vitro* studies showed the activation of some of the molecules involved in the innate immune response. However, they did not lead to the identification of a receptor that recognises gluten nor that demonstrates that gluten, rather than another antigen, is in fact responsible for activating the innate immune response in CD.

In chapter 4, we described the results from gene and protein expression experiments on biopsy material from coeliac patients and controls that provide evidence for activation of the innate immune response. This data showed an important role for the HLA-E molecules at the epithelial layer, genes associated with the NF-κB pathway, and a raised level of neutrophils in the lamina propria of coeliac individuals in total remission (M0 stage). These observations did not point to a receptor or specific stimulus responsible for activating the innate immune response in CD. However, the infiltration of neutrophils at the M0 stage (chapter 4) (Fig. 1b) suggests that a permanent activation of the innate immune response in the coeliac mucosa is not due to gluten, but rather reflects an impaired permeability of the intestinal barrier. This is in line with the recent observation of a primary genetic variant in the *MYO9B* gene in coeliac patients, which might also undermine the permeability of the epithelial barrier [17]. It is tempting to speculate that during the progression of the disease, this activated innate reaction could be exacerbated by IL-15 since it activates the NF-κB pathway in neutrophils, increases their chemotaxis to inflamed sites and their function, and delays their apoptosis [18]. These

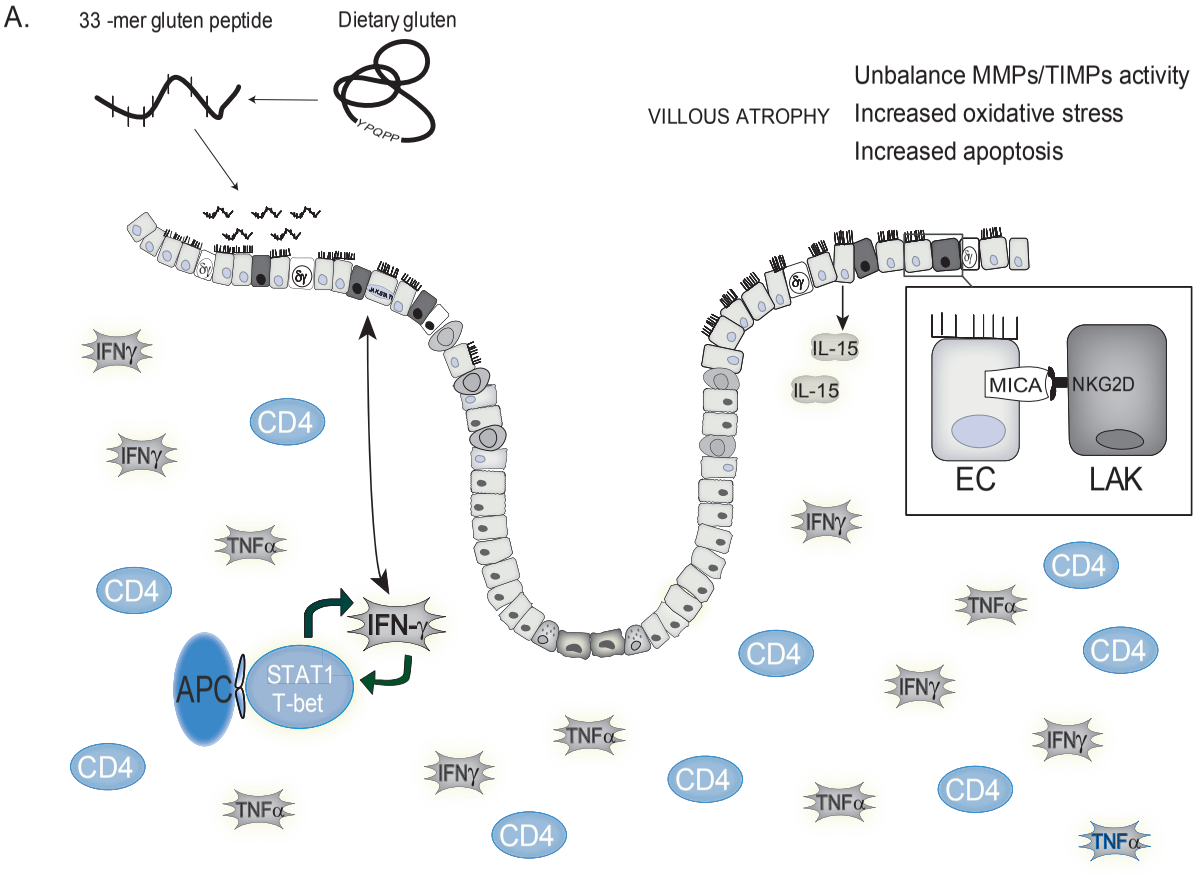
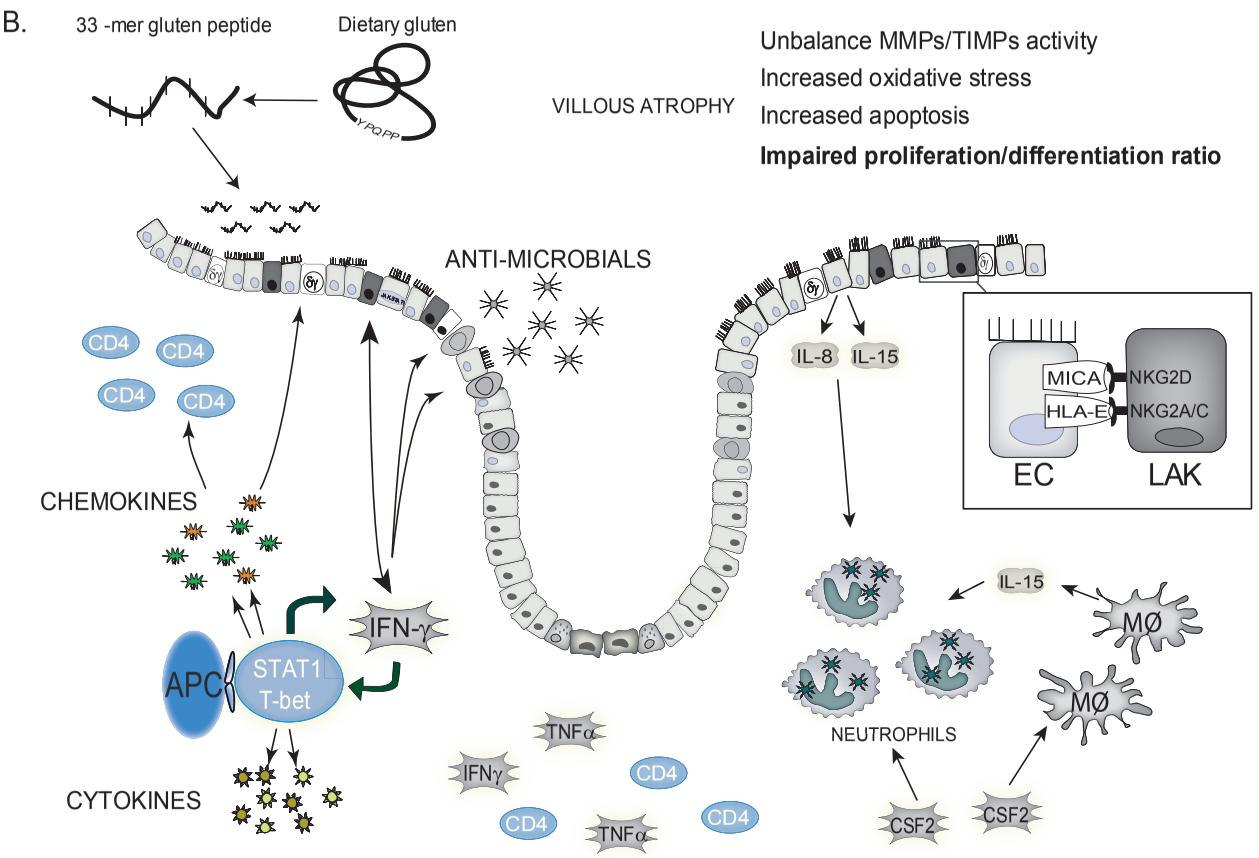


Figure 1. Global overview of the molecular processes in the mucosa of coeliac disease individuals. a) The figure depicts the adaptive Th1 immune response mediated by HLA-DQ2/DQ8 antigen presenting cells (APCs), which present immuno-dominant gluten peptides to a repertoire of αβ CD4+ T cells that then become activated. It also shows the increased expression of IFN-γ and TNF-α in the lamina propria. The innate immune response is activated at the epithelial layer by non-immuno-dominant gluten peptides and IL-15 through intraepithelial lymphocytes (LAK) that express the receptor CD94/NKG2D and recognise MHC class I polypeptide-related sequence A (MICA) molecules expressed on stressed epithelial cells (EC).



b) The figure illustrates the postulated functions of IFN-γ in maintaining a chronic Th1 inflammatory response at the epithelial layer and the lamina propria. This might take place by activating the JAK-STAT1 pathway in the αβ CD4+ T cells and the epithelial cells, and by secreting chemokines and cytokines that attract αβ CD4+ T cells and γδ CD8+ T cells into the mucosa. Our results indicated that IFN-γ might also regulate the expression of innate-related genes with anti-microbial function and non-classical HLA type I molecules, such as HLA-E that can bind to the receptor CD94/NKG2A/C and reduce the innate immune response at the epithelial layer. In addition, the activation of the innate response at the lamina propria is recognised by the increased expression of IL-8 and colony stimulating factor 2 (CSF2) that attract macrophages and neutrophils, whose function can be enhanced by IL-15.

activated neutrophils could participate in the immune response phagocytating death cells or killing bacteria and viruses that reach the mucosa of CD patients.

Future studies should focus on identifying a coeliac-specific receptor that activates the innate system, its primary contribution to the disease, and on which antigen is responsible for its activation. This would elucidate whether an innate reaction is essential in the immunopathogenesis of the disease, or whether it only occurs in a subset of the coeliac population. An example of such a subset could be the individuals that develop enteropathy-associated T-cell lymphomas. This is a specific complication of CD in which IL-15 mediates the proliferation of the intraepithelial lymphocytes that then become malignant [19]. Searching for this receptor in a subset of the coeliac population might be a fruitful strategy to determine its identity.

The participation of a Th1 adaptive immune response is an established element in the pathogenesis of CD. It is well understood that HLA-DQ2/DQ8 molecules and the high levels of IFN-γ activate αβ CD4+ T-cells. As explained in chapter 1, the proposed function of the adaptive immune response in the pathogenesis of CD was mediating villous atrophy [20]. However, there was no consensus on the molecular mechanisms leading to villous atrophy, and three different hypotheses were proposed (chapters 1 and 2). The first hypothesis postulated that the adaptive immune response drives villous atrophy through the activation of αβ CD8+ T cells expressing FASL and interacting with FAS+ enterocytes. A second mechanism was proposed for activation of the immune response, increasing the activity of metalloproteinases and consequently destroying structural elements of the lamina propria. The third hypothesis proposed that the activated immune response and the activity of some transcription factors such as STAT-1 increase the oxidative stress in the lamina propria. We looked at the transcriptional activity of the genes in the light of these three hypotheses.

We observed an expected increased expression of IFNG and downstream cytokines and chemokines of the JAK-STAT1 pathway. Our results suggested that IFN-γ functions in CD by maintaining a chronic inflammatory response at the epithelial layer and the lamina propria. This is presumably done by attracting γδ CD8+ - and αβ CD4+ T-cells, and activating the JAK-STAT1 pathway at the lamina propria and the epithelial layer (chapter 3). We noted that IFN-γ might also regulate the expression of innate-related genes, such as *RNasel* which induces apoptosis of infected cells [21] (Fig. 1b). It was surprising that none of the proposed pathways for mediating villous atrophy showed significant differences in the expression studies. Furthermore, immunohistochemistry studies with activated caspase-3 did not show more positive staining on the epithelial layer in CD patients than in controls, and inducible nitric oxide synthase (iNOS) showed an increase inmmunoreactivity in CD patients only at the lamina propria (unpublished data, Fig. 2). Another mechanism might therefore be responsible for flattening the mucosa.

An interesting question that needs further research is how the adaptive and the innate immune response cross talk in CD. In the gut-associated lymphatic tissue, the innate and the adaptive responses cooperate to maintain a balanced immune response. Candidate cells holding this cross-talking are epithelial cells and dendritic cells. Epithelial cells express Toll-like receptors, produce anti-microbial peptides, and recruit cells of the immune system. Dendritic cells are able to initiate the adaptive immune response by stimulating naive T-cells

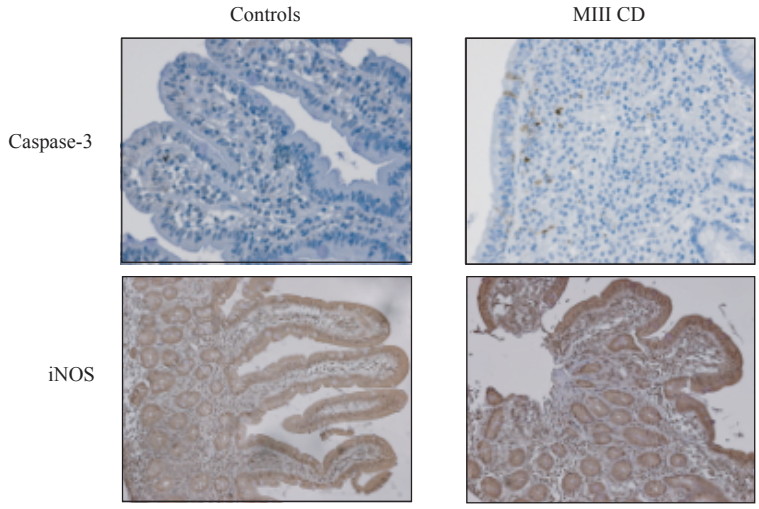


Figure 2. Immunohistochemistry studies on pathways postulated to drive villous atrophy in coeliac disease. The presence of apoptosis in the villi of four CD MIII and four controls as described in chapter 4 was assessed by anti-caspase3 antibody, which detects the activated form of the apoptotic-related caspase-3. No differences in inmmunoreactivity were detected between patients and controls. Inducible nitric oxide synthase (iNOS) staining in the mucosa of CD patients showed increased inmmunoreactivity in the lamina propria of the MIII CD patients compared to controls. No differences were found at the epithelial layer.

into Th1, Th2 or T regulatory cells, and sample the bacteria present in the lumen through dendrites that open up intercellular junctions of the epithelial barrier. In CD, it is possible to assume that gluten or another antigen could stimulate epithelial cells to produce chemokines, cytokines and anti-microbial peptides that attract lymphocytes, neutrophils, macrophages and immature dendritic cells into the mucosa. Upon maturation, the dendritic cells would recognise gluten in an inflammatory milieu and present it to naive T-cells as a danger signal in the mesenteric nodes. Determining how the epithelial cells let gluten reach the lamina propria, how their activation shapes the immune response, and the attributes of the dendritic cells in the lamina propria, Peyer's patches and mesenteric nodes of CD patients will help to identify how the innate- and the adaptive immune responses join forces in CD.

Villous atrophy: a new pathway

The small intestine is a dynamic and rapidly renovated organ organised in crypt-villous units. Crypts contains a limited number of stem cells that continuously give rise to daughter cells that further divide and receive signals from their environment to commit to a particular cell fate, to cell cycle arrest, or to achieve final differentiation into Goblet cells, enteroendocrine cells or enterocytes while they migrate towards the lumen. Once differentiated, these cells enter into the villous where they perform their function and are then shed at the top of the villous after five days. In summary, the maintenance of the structure of the crypt villous units is maintained by: i) cell production in the crypts to compensate cell shedding at the top of the villous, ii) continuous ascending migration of the dividing as well as the differentiated cells, and iii) non-cell autonomous processes that determine the cell proliferation/differentiation stages of these units [14, 15]. Thus, it is easy to see that in CD any alteration in the balance of

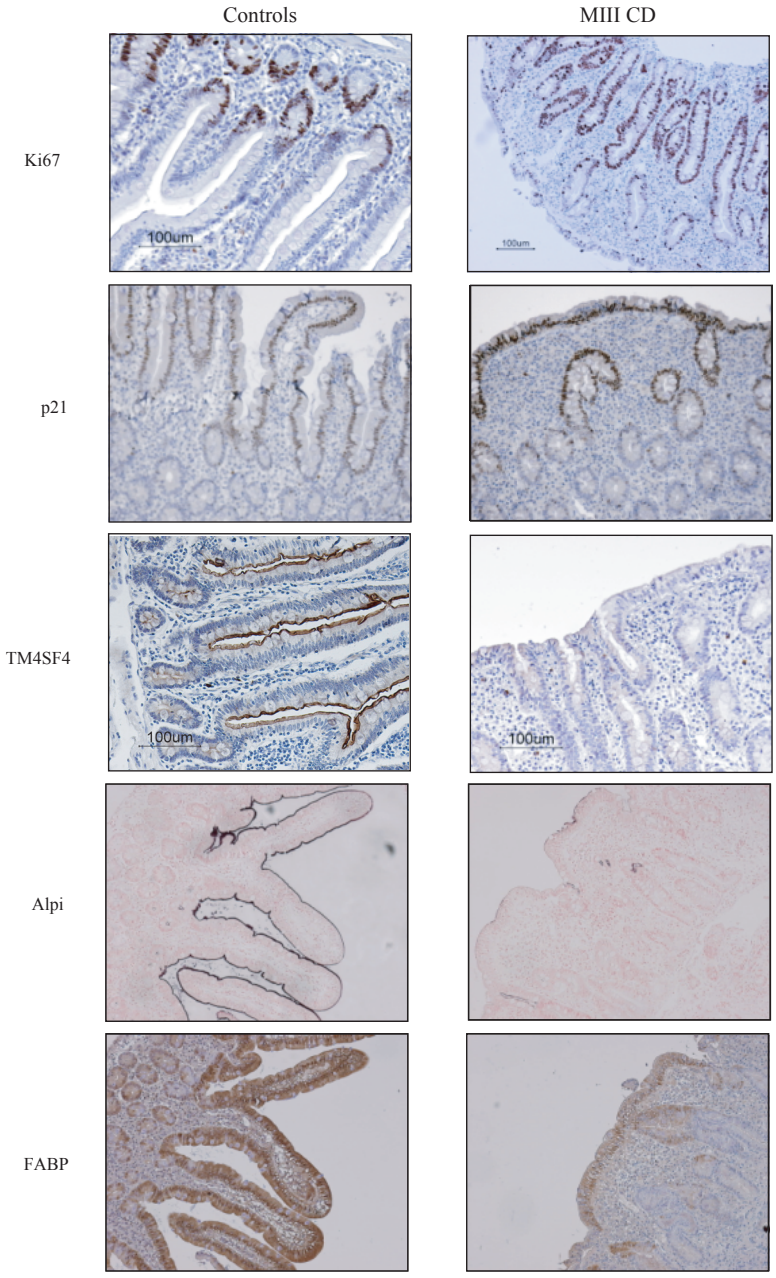


Figure 3. Immunohistochemistry studies on proliferation and differentiation pathways in coeliac disease. The proliferation state of the cells of the crypts in CD patients and controls was evaluated in four CD MIII and four controls as described in chapter 4 with the ki-67 marker, which showed increased cellularity of the crypts in patients compared to controls. To further examine the differentiation state of the cells of the villi, the marker p21, which shows cell arrest, was tested but showed no differences between MIII CD patients and controls. However, the markers for terminal differentiation tetraspanin family member 4 (TM4SF4) and alkaline phosphatase (Alpi) did show decreased and patchy staining in biopsies from MIII patients compared to controls, although the fatty acid-binding protein (FABP) showed no differences in inmmunoreactivity.

these processes might lead to crypt hyperplasia and villous atrophy.

Identifying and understanding the molecular mechanisms that drive villous atrophy are an important issue in CD because most of the clinical symptoms seen in coeliac patients are due to the flattening of the mucosa. As described in chapter 2, three alternative hypotheses on how the immune response triggers villous atrophy have been proposed: increased apoptotic activity in the epithelial layer, increased activity of MMPs, and enhanced oxidative stress [16]. However, neither our expression studies at mRNA nor at protein level showed differences in expression that would indicate the participation of any of these pathways in the villous atrophy (unpublished data, Fig. 2).

Comparing the expression profiles of biopsies from MIII patients to those from controls suggested that villous atrophy was due to a deficiency of cells entering terminal differentiation and exiting the crypts (chapter 2) (Fig. 1b). This was further substantiated by comparing MIII biopsies from patients ingesting or refraining from gluten. This comparison showed that the first molecular event towards normalisation was decreased mitotic activity of the crypts, suggesting that cells stop proliferating and can therefore enter differentiation (chapter 4) [22]. To further determine the differentiation stage of the cells in the villous, we conducted immunohistochemistry studies on biopsies from MIII patients and controls. These experiments showed more proliferative cells in the MIII crypt compartment with the ki67 proliferation marker (as expected), but no differences with the cell cycle inhibitor, cyclin p21, indicating that cells of atrophic villi could enter cell cycle arrest. However, terminal differentiation seemed to be impaired as its markers, tetraspanin family member 4 (TM4SF4), and alkaline phosphatase (Alpi) showed decreased or patchy staining in biopsies from MIII patients compared to those from control. A third marker of terminal differentiation, the fatty acid-binding protein (FABP) did not show differences in inmmunoreactivity (Fig. 3, unpublished data). The transcription of several brush border enzymes, such as the fructose-1,6-bisphosphatase (FBP) or aminopeptidase N (ANPEP), was upregulated during the restoration of the crypt villous units when on a gluten-free diet (unpublished data).

To provide support for this hypothesis, in which crypt hyperplasia and villous atrophy are due to an impaired ratio between the proliferation and terminal differentiation of the enterocytes, future research should include: expression studies on single cell types to detect which genes and pathways are disturbed and look at their functional follow-up with cyto-histological techniques, *in vitro* systems such as tandem affinity purification (TAP) and RNAi, and *in vivo* models and genetic association studies to show the molecular signature, the exact function of the cells at specific locations in the crypt villous units, and their primary effect in CD. Knowing which molecules and signalling cascades are impaired, and in which cells, will help determine how the proliferation/differentiation processes might be impaired in the disease. This would lead to answers to more specific questions: i) whether cells do not enter irreversible growth arrest but execute some differentiation processes, ii) whether there is a distortion in the number of each cell type located at the atrophic villi, and iii) whether upwards migration of the differentiated cells is impaired.

Identification of markers for the disease progression

The coeliac mucosa show a characteristic histological development that can be used to

classify the disease into three different stages and to diagnose the disease. The gold standard for diagnosing the disease is a biopsy with a Marsh III lesion while the patient is still ingesting gluten, and rapid clinical improvement when on a gluten-free diet (ESPGHAN criteria) [23]. These criteria do not consider individuals with a biopsy showing an MI or MII lesion as CD patients, and there is no recommendation on how to manage this situation, nor reference to histological healing after treatment. Studies on the histological improvement in CD individuals show that, despite their rapid clinical improvement, a complete histological recovery can take up to two years in adults, depending on the degree of severity at the time of diagnosis, and recovery is not complete in most cases [24, 25]. These observations also indicate that clinical normalisation is not directly reflected in a restored histological normality. In our study we aimed to identify genes that could be used as markers for each of the histological stages that would facilitate the histological classification and the diagnosis. As explained in chapter 4, our expression studies on biopsies of the sequential Marsh types showed no differences in expression between consecutive stages, while plotting the expression levels for each of the histological categories revealed two groups of genes with gradually changing profiles in opposite directions (chapter 4, Fig. 1). As can be appreciated from Fig. 1, this work did not lead to identifying any classifiers. The results indicated that, despite the morphological differences, MI and MII individuals cannot be distinguished at a molecular level from MIII individuals and, based on our results, we suggest that they should be classified as CD patients. This is further supported by studies in which a subset of MI and MII individuals developed an MIII lesion after a gluten challenge [26]; more studies to better define these patients need to be conducted. One possibility lies in population-based studies to uncover the incidence of complications often seen in untreated MIII CD patients, such as lymphomas or osteoporosis in MI and MII individuals. An increased incidence of such complications in MI and MII individuals could be explained if they are considered as untreated CD patients. A second option is to identify causative genes and use them as diagnostic coeliac markers; these markers are traditionally identified in linkage and association studies. In our study, we reasoned that comparing the expression profiles of biopsies from patients in total remission (M0) to control biopsies should point towards any differentially expressed genes that might be causally related to CD. This experiment revealed two genes: the family with sequence similarity 12, member A (*FAM12A*) and the T cell receptor associated transmembrane adaptor 1 (*TRAT1*). It is now necessary to determine whether they have a primary association to CD, or are secondary markers specific for the disease. In either case, these findings could have direct repercussions in day-to-day diagnostic pathology since these genes could be used as markers to discriminate between non-coeliac individuals with a MI/MII-like biopsy due to another enteropathology, and true MI/MII coeliac patients. It would also be interesting to determine whether their level of expression in blood corresponds to measurements in the gut, and whether they could be used as markers for screening high-risk populations and identifying silent individuals (with no clinical symptoms).

Identification of putative primary disease genes

Two approaches have been used to identify candidate genes in CD: genome-wide screens

and genetic association studies on candidate genes. To date, several genome screens have been performed in a number of mainly European populations [27]. However, despite their significant findings, the studies have not been able to replicate each other's results, except for the chromosome 6p21 region containing the *HLA-DQ2/DQ8* genes. Only the genome-wide screen performed in the Dutch coeliac population has been able to identify the gene responsible for the significant linkage [17]. These observations could be due to under-power, founder effects, incorrect inclusion criteria, or the fact that CD is a multifactorial disorder and therefore not all the genes make an equal contribution in the different populations. Thus, the challenge of looking for causative genes in CD needs some alternative strategies.

One possibility of identifying primary genes for common multifactorial diseases such as CD is to perform genome-wide association studies. This is now possible because tag SNPs from the completed HapMap project can be used [28, 29]. This type of study is statistically more powerful since we only need to collect single cases instead of sibpairs, and it is more effective since the genotypes determined are based on the knowledge of genome-wide linkage disequilibrium. One successful example of the tag SNP approach is the follow-up association study carried out using 359 tag SNPs on the 99% confident interval on chromosome 19 in two cohorts of 216 and 247 Dutch coeliac patients, and 216 and 470 controls, respectively. This study was solely based on the HapMap information and led to the identification of the first non-HLA gene responsible for CD, the myosin IXB gene (*MYO9B*) [17]. Further haplotype analysis showed that individuals heterozygous for the A allele of SNP rs2305764 have an elevated risk of 1.66, while homozygous individuals have a risk of developing CD that increases to 2.27, with population attributable risks of 25% and 28%, respectively.

Another strategy uses bioinformatic tools to integrate different sources of genomic data. We used bioinformatic tools developed in-house to integrate genetic data with genome-wide gene expression data to identify differentially expressed genes in linkage areas that could then be considered as putative primary genes [30]. This led to the identification of two candidates for genetic association studies: prolyl endopeptidases (*PREP*) and the signal transducer and activator of transcription 1-alpha/beta (*STAT1*) [31, 32]. Although both these genes were convincing candidate genes (PREP is an enzyme able to digest gluten, and STAT1 modulates the activity of IFN-γ), neither gene showed positive association to the Dutch CD population (chapters 5 and 6).

An alternative approach to prioritizing putative causative genes in CD is to combine genome-wide expression and genotype data to identify those genes whose expression is strongly correlated with a specific genotype (a so-called expression QTL or eQTL). Such eQTL regions may contain causative genotypic polymorphisms providing a novel method, known as genetical genomics, for identifying candidate genes [33]. The most obvious application of combining genome-wide expression and genotype data through genetical genomics in CD is the identification of new chromosomal regions that might hold causative genes or key elements regulating the expression of other genes. In this latter situation, defining such "regulators" might help create a "pathogenic coeliac network". Moreover, combining eQTLs from different autoimmune disorders could help identify common regulatory mechanisms, while stratifying the genetic data might unravel interacting loci and the downstream effects

of some known genetic factors.

Looking into the future

One of the major challenges in this study was to interpret the data from such a heterogeneous tissue type as the coeliac biopsies. These biopsies consist of cells whose physiological function may not be altered at all by the disease, cells that do not have a primary role in the disease but whose behaviour is altered as a consequence of the changed micro-environment, and cells which are the cause of the disease. Consequently, our expression studies reflected the expression signature of physiological processes of the intestine, and of the primary and secondary effects of gluten ingestion by genetically predisposed individuals. In this situation, the true disease signal was diluted into the background by so much noise and we could only detect the tip of the ice-berg. Moreover, our limited knowledge about the differentially expressed genes, the gut biology and the underlying molecular processes in CD did not enable us to identify specific genes expressed in certain cell types, but only the relevant pathways. We had to distinguish which signature profiles were CD-specific, and which were merely secondary or physiological.

One way to circumvent this difficulty in CD would be to determine the signals and expression profiles from single cell type populations. This can be achieved using techniques that allow the retrieval of individual or specific groups of cells, such as isolation of epithelial cell layer, γδ and αβ T-cells, and laser capture microdissection microscopy (LCM). This type of experiment will enable us to assign signals to specific cell types, to identify signals from cells present in small numbers, and to identify relevant genes expressed in cell populations that we suspect are important in the pathogenesis of CD. Hence, this approach could have a direct application in investigating the crypt cells at different heights in the crypts and the pericryptal stromal cells. Determining their expression profiles will help define the molecular pathways that underlie crypt hyperplasia and show whether stromal cells can send inflammatory signals to cells in the crypts to influence their proliferation/differentiation status. It would be reasonable to complement this experiment by exploring the epithelial cell layer to further investigate whether the intermingled proliferation and differentiation processes do indeed lead to crypt hyperplasia and villous atrophy, as we proposed in this research. Another very interesting possibility in using LCM would be to compare the expression profiles of the cells that express *MYO9B* from individuals with different genetic susceptibilities i.e. high risk versus low risk. This could help elucidate the function of MYO9B in the context of cell signalling and validate MYO9B's involvement in the impairment of the intestinal barrier in CD.

However, the findings of these expression studies would only show the regulation at mRNA level and, since the function and regulation of many genes occur at a protein level, functional experiments that look into post-translational modifications (e.g. phosphorylation, glycosylation or acetylation) at the protein level would still be necessary. For example, immuno¬histochemistry, fluorescent microscopy or in situ hybridisation experiments could determine where MYO9B, other key genes, and the differentially expressed genes related to proliferation/differentiation pathways are located in the intestine. Further mechanistic insights into the proposed hypothesis could be achieved by studying *in vitro* models, such

as cellular systems (even under different culture conditions) and the knock-down and over-expression of target genes to identify signalling pathways and determine *in vitro* phenotypes. Animal models could be used to test the generated hypotheses and determine the function of genes in a complex *in vivo* situation. Functional approaches and especially proteomic approaches with 2D gels and mass spectrometry could be undertaken in parallel to genome-wide expression and genetic studies. These types of experiments would help in analysing the expression profiles of heterogeneous tissues by integrating data with bioinformatics tools.

Conclusions

Our expression studies on the biopsies of CD patients and controls have identified new pieces in the puzzle of coeliac pathogenesis, and have pinpointed areas of the puzzle we should look at. A better understanding of the molecular mechanisms of the pathogenesis of CD will indicate new targets for developing treatments and diagnostic tools that could be used in day-to-day clinical work and in screening populations to identify silent patients or to characterise subsets of coeliac patients at risk of developing severe complications.

References

1. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 1992; 102: 330-54.

2. Louka AS and Sollid LM. HLA in coeliac disease: unravelling the complex genetics of a complex disorder. *Tissue Antigens* 2003; 61: 105-17.

3. Sollid LM. Molecular basis of celiac disease. *Annu Rev Immunol* 2000; 18: 53-81.

4. Kontakou M, Sturgess RP, Przemioslo RT, Limb GA, Nelufer JM and Ciclitira PJ. Detection of interferon gamma mRNA in the mucosa of patients with coeliac disease by in situ hybridisation. *Gut* 1994; 35: 1037-41.

5. 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* 2004; 23: 183-90.

6. Westerholm-Ormio M, Garioch J, Ketola I and Savilahti E. Inflammatory cytokines in small intestinal mucosa of patients with potential coeliac disease. *Clin Exp Immunol* 2002; 128: 94-101.

7. Veres G, Westerholm-Ormio M, Kokkonen J, Arato A and Savilahti E. Cytokines and adhesion molecules in duodenal mucosa of children with delayed-type food allergy. *J Pediatr Gastroenterol Nutr* 2003; 37: 27-34.

8. Lahat N, Shapiro S, Karban A, Gerstein R, Kinarty A and Lerner A. Cytokine profile in coeliac disease. *Scand J Immunol* 1999; 49: 441-6.

9. Breese EJ, Kumar P, Farthing MJ and MacDonald TT. Interleukin-2 and interferon-gamma producing cells in the lamina propria in celiac disease. *Dig Dis Sci* 1994; 39: 2243.

10. Seegers D, Borm ME, van Belzen MJ, Mulder CJ, Bailing J, Crusius JB, Meijer JW, Wijmenga C, Pena AS and Bouma G. IL12B and IRF1 gene polymorphisms and susceptibility to celiac disease. *Eur J Immunogenet* 2003; 30: 421-5.

11. Salvati VM, MacDonald TT, Bajaj-Elliott M, Borrelli M, Staiano A, Auricchio S, Troncone R and Monteleone G. Interleukin 18 and associated markers of T helper cell type 1 activity in coeliac disease. *Gut* 2002; 50: 186-90.

12. Polvi A, Maki M, Collin P and Partanen J. TNF microsatellite alleles a2 and b3 are not primarily associated with celiac disease in the Finnish population. *Tissue Antigens* 1998; 51: 553-5.

13. Maiuri L, Ciacci C, Ricciardelli I, Vacca L, Raia V, Auricchio S, Picard J, Osman M, Quaratino S and Londei M. Association between innate response to gliadin and activation of pathogenic T cells in coeliac disease. *Lancet* 2003; 362: 30-7.

14. Meresse B, Chen Z, Ciszewski C, Tretiakova M, Bhagat G, Krausz TN, Raulet DH, Lanier LL, Groh V, Spies T, Ebert EC, Green PH and Jabri B. 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.

15. Hue S, Mention JJ, Monteiro RC, Zhang S, Cellier C, Schmitz J, Verkarre V, Fodil N, Bahram S, Cerf-Bensussan N and Caillat-Zucman S. A direct role for NKG2D/MICA interaction in villous atrophy during celiac disease. *Immunity* 2004; 21: 367-77.

16. Diosdado B, van Oort E and Wijmenga C. "Coelionomics": towards understanding the molecular pathology of coeliac disease. *Clin Chem Lab Med* 2005; 43: 685-95.

17. 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 EV, Wapenaar MC, Koeleman BP and 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.

18. Cassatella MA and McDonald PP. Interleukin-15 and its impact on neutrophil function. *Curr Opin Hematol* 2000; 7: 174-7.

19. Mention JJ, Ben Ahmed M, Begue B, Barbe U, Verkarre V, Asnafi V, Colombel JF, Cugnenc PH, Ruemmele FM, McIntyre E, Brousse N, Cellier C and Cerf-Bensussan N. Interleukin 15: a key to disrupted intraepithelial lymphocyte homeostasis and lymphomagenesis in celiac disease. *Gastroenterology* 2003; 125: 730-45.

20. Diosdado B and 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.

21. Li G, Xiang Y, Sabapathy K and Silverman RH. An apoptotic signaling pathway in the interferon antiviral response mediated by RNase L and c-Jun NH2-terminal kinase. *J Biol Chem* 2004; 279: 1123-31.

22. 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* 2004; 53: 944-51.

23. 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.

24. Wahab PJ, Meijer JW and 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.

25. Lee SK, Lo W, Memeo L, Rotterdam H and Green PH. Duodenal histology in patients with celiac disease after treatment with a gluten-free diet. *Gastrointest Endosc* 2003; 57: 187-91.

26. Wahab PJ, Crusius JB, Meijer JW and Mulder CJ. Gluten challenge in borderline gluten-sensitive enteropathy. *Am J Gastroenterol* 2001; 96: 1464-9.

27. van Heel DA, Hunt K, Greco L, Wijmenga C. Genetics in coeliac disease. Best Pract res clin Gastroenterol 2005;19:323-39

28. Hirschhorn JN, Daly MJ. Genome-wise association studies for common diseases and complex traits. Nat Rev. Genet 2005;6:95-108.

29. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IS, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto J, Cough SC, Clayton DG, Todd JA. Haplotype mapping for the identification of common disease genes. Nat Gent 2001;29:233-7.

30. 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; 12: 633-8.

31. Diosdado B, Stepniak DT, Monsuur AJ, Franke L, Wapenaar MC, Mearin ML, Koning F and 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.

32. Diosdado BMA, Mearin ML, Mulder CCJ, Wijmenga C. The down-stream modulator of interferon-gamma, STAT1 is not genetically associated to the Dutch celiac disease population.

33. Doerge RW. Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet* 2002; 3: 43-52.

34. Duggan DJ, Bittner M, Chen Y, Meltzer P and Trent JM. Expression profiling using cDNA microarrays. *Nat Genet* 1999; 21: 10-4.

35. Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, Sallan SE, Lander ES, Golub TR and Korsmeyer SJ. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* 2002; 30: 41-7.

36. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R and Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415: 530-6.

37. Gerhold DL, Jensen RV and Gullans SR. Better therapeutics through microarrays. *Nat Genet* 2002; 32 Suppl: 547-51.

38. Lockhart DJ and Winzeler EA. Genomics, gene expression and DNA arrays. *Nature* 2000; 405: 827-36.

SUMMARY

Coeliac disease (CD) is a genetic disorder in which environmental factors and multiple genes contribute to its development. The main environmental factor that triggers CD is a protein called gluten. This protein is present in common cereals such as wheat, barley and rye, and allows us to process their derivate products such as bread or pasta. The only known genetic factors for CD are the genes located in chromosome 6 that encode for the HLA-DQ2/DQ8 molecules, and the *MYO9B* gene located on chromosome 19. Clinical symptoms include a wide rage of features that vary from typical diarrhea and malabsorption (or failure to thrive in children), to fatigue, anaemia, weight loss or no symptoms at all. Currently, the only therapy consists of a strict, life-long, gluten-free diet.

The disease starts when a susceptible individual ingests gluten and this enters his/her body through the mucosa of the small intestine. There, gluten is able to activate an abnormal immune response that is going to lead to consecutive, well-established changes of the intestinal mucosa. These are divided into three stages: lymphocytosis (Marsh I stage), crypt hyperplasia (Marsh II) and villous atrophy (Marsh III). Over the past few years, our knowledge about the pathogenesis of CD has increased substantially. We have long known that gluten could activate both adaptive and innate responses. However, further characterization of the immune response and the mechanisms behind tissue remodelling are still poorly understood (chapters 1 and 2). The research described in this thesis provides new insights into the molecular mechanisms of both the adaptive and innate immune response, and the molecular mechanisms that drive villous atrophy. Consequently, uncovering new aspects of the patho-physiology of the disease has directed our search for candidate genes that predispose to CD.

In Chapter 3 we describe the transcriptional changes that occur in the intestine of CD individuals who show villous atrophy (Marsh III) compared to healthy controls. We compared the expression profiles of biopsies from 15 well-characterized Marsh III individuals and seven controls in a microarray experiment that encompassed 19,500 genes. We reported a set of 109 genes involved in the long-term tissue destruction. Many of these genes had a function in proliferation and differentiation pathways and might be involved in the homeostasis of the crypt villi. This was further supported by comparing the expression profiles from Marsh III patients refraining from gluten to Marsh III patients ingesting gluten. Our analysis yielded a second set of 120 genes that revealed a decreased mitotic activity in Marsh III patients after gluten withdrawal. Based on these results we proposed that the separate processes of crypt hyperplasia and villous atrophy are molecularly intertwined and we hypothesized that a deficiency of the cells in the crypts in completing their terminal differentiation program and migrating towards the villous was responsible for the formation of hyperplasic crypts and the subsequent villous atrophy.

A second microarray experiment was carried out to specifically determine the transcriptional activity of the genes related to the immune response during mucosal renewal. In chapter 4 we describe how a total of 66 biopsies from CD individuals in various stages of remission and from controls enabled us to set up a virtual time course experiment. This allowed us to determine immune-related differentially expressed genes and the dynamics of their pattern of expression during tissue regeneration. The results revealed that gene expression profiles from the differentially expressed genes showed a gradual increase or decrease from the moment

the treatment was started to the point of full recovery. It was interesting that this observation did not correlate with the discrete histological (Marsh) stages used in diagnosing the disease. Thus, despite the morphological differences, Marsh I and Marsh II individuals cannot be distinguished at a molecular level from Marsh III patients. Hence, we proposed that all patients should be considered "coeliacs" as long as they respond to a gluten-free diet.

The classification of the genes within the adaptive and innate pathways showed a comparable contribution of both pathways in response to gluten withdrawal. Our results suggested that interferon gamma is able to maintain the chronicity of the inflammatory response by inducing the transcription of chemokines and cytokines that attract and further activate the Th1 immune response through the JAK-STAT1 pathway. In relation to the innate immune response, our expression studies showed several indicators of neutrophils being recruited into the lamina propria of CD patients. This observation was further confirmed by immunohistochemistry using neutrophil-specific antibodies in both Marsh III and completely recovered (Marsh 0) CD individuals. This result suggests that CD patients have a permanent state of activation of the innate immune response in the lamina propria, regardless of whether they have histological damage or exposure to dietary gluten. We speculated that the enhanced neutrophil infiltration might reflect an underlying effect of the *MYO9B* gene. The MYO9B might be involved in the impairment of the intestinal barrier and, hence, lead to an increased permeability as seen in CD individuals.

The integration of the gene expression data with known linkage data, combined with knowledge on the pathogenesis of the disease, led us to two functional candidate genes that were subsequently tested in genetic association studies described in chapters 5 and 6. The prolyl-endopeptidase (*PREP*) enzyme, located under the 6q21-22-linkage peak, is able to cleave proline-rich gluten peptides. We hypothesized that an altered PREP activity could be responsible for an inefficient breakdown of gluten peptides, which could result in the accumulation of immunostimulatory gluten peptides and thereby play a role in the breakdown of tolerance to gluten. To investigate the role of the *PREP* gene as a primary factor in CD, we conducted sequence analyses, genetic association studies, and determined the PREP enzyme activity in biopsies from CD patients and controls. Our results, described in chapter 5, from the genetic association studies and the activity determinations indicated that PREP is not a causative gene in the Dutch CD population.

Chapter 6 describes the association studies performed on the positional and functional candidate signal transducer and activator of transcription 1 (*STAT-1*) gene, located under the 2q linkage peak. This gene encodes for one of the main down-stream modulators of interferon gamma. We speculated that any genetic change that alters its transcriptional activity could explain the detrimental immune response seen in CD. Again, no evidence for positive association was observed in the Dutch CD population.

The gene expression studies performed in this project on the biopsies of CD patients and controls have led to the identification of novel pathways involved in the pathogenesis of CD. This new knowledge will help the search for susceptibility genes in CD. Moreover, the integration of the genetic and genomic information will be key to developing new diagnostic tools for identifying individuals at risk, and to finding new avenues for therapeutic intervention in CD.

SUMARIO

La enfermedad celiaca es un desorden de carácter genético cuya etiología se atribuye tanto a múltiples factores genéticos como a factores medioambientales. El principal factor medioambiental conocido que desencadena la enfermedad es una proteína llamada gluten que se encuentra en cereales comunes como el trigo, la cebada y el centeno. Esta proteína permite la germinación de las semillas y gracias a ella se pueden elaborar productos derivados de sus harinas como el pan o la pasta. Los desencadenantes genéticos de la enfermedad conocidos hasta el momento son los genes localizados en el cromosoma 6, responsables de la codificación de las moléculas HLA-DQ2/-DQ8, y el gen que codifica la miosina 9B (MYO9B) y que se encuentra en el cromosoma 19. Las manifestaciones clínicas de esta enfermedad incluyen una amplia variedad de síntomas que van desde diarrea, mala absorción o retraso en el crecimiento ponderal del niño, hasta fatiga, anemia, pérdida de peso o formas totalmente asintomáticas. Hoy en día, el único tratamiento consiste en una dieta permanente sin gluten.

La enfermedad celiaca comienza cuando un individuo con la correcta predisposición genética ingiere gluten y éste entra en su organismo a través de la mucosa de su intestino delgado. Allí el gluten activa una respuesta inmune anormal que va a desencadenar una serie de cambios histológicos conocidos y que consisten en linfocitosis (estadio I; Marsh I), hiperplasia de las criptas (estadio II; Marsh II) y atrofia de las vellosidades intestinales (estadio III; Marsh III). Es sabido que estos cambios ocurren siempre en este orden y según criterios diagnósticos internacionales, sólo individuos que presentan este tercer estadio son considerados enfermos celiacos. En los últimos años, nuestro conocimiento sobre la patogenia de la celiaca ha aumentado considerablemente. Sabemos que el gluten es capaz de activar tanto la respuesta inmune adaptativa como la innata. A pesar de ésto, los mecanismos intrínsecos de esta respuesta inmune y los mecanismos moleculares responsables del daño en la mucosa intestinal no son bien conocidos (capítulo 1 y 2). El trabajo descrito en esta tesis revela nuevos hallazgos sobre los mecanismos moleculares que dirigen la respuesta inmune adaptativa e innata, así como las alteraciones de la mucosa. Como consecuencia, estos nuevos conceptos sobre la pato-fisiología de la enfermedad celiaca enfocan la búsqueda de genes candidatos que predisponen a desarrollar esta enfermedad.

El capítulo 3 describe los cambios transcripcionales que ocurren en el intestino de los pacientes celiacos con atrofia intestinal (estadio III; Marsh III) cuando se les compara con controles. Para ello comparamos los patrones de expresión de 15 biopsias de enfermos celiacos en el estadio III de la enfermedad y 7 controles en un microarray con 19,500 genes. El primer análisis de estos experimentos reveló un primer grupo de 109 genes implicados en la destrucción del intestino a largo plazo. Una gran parte de estos genes se encuentran relacionados con los procesos de proliferación y diferenciación que tienen lugar en las criptas y las vellosidades y que contribuyen al mantenimiento de las mismas. Este hallazgo fue corroborado por un segundo análisis en el que comparamos los patrones de expresión de pacientes que presentan atrofia intestinal ocasionada por la presencia de gluten en la dieta, con los de pacientes que aún presentan alteraciones histológicas a pesar de seguir una dieta sin gluten, pero con mejoría clínica. De este modo, encontramos un grupo de 120 genes que revelan un descenso de la actividad mitótica en pacientes que siguen una dieta sin gluten.

Basándonos en estos resultados, proponemos que los procesos independientes de hiperplasia de las criptas y atrofia de las vellosidades están molecularmente entrelazados y planteamos la hipótesis de que una deficiencia en la finalización de la diferenciación terminal de las células epiteliales y su migración es la responsable de la formación de criptas hiperplásticas y, consecuentemente, de la atrofia vellositaria. A continuación, llevamos a cabo un segundo microarray con el fin de determinar los cambios transcripcionales de genes relacionados con la respuesta inmune durante la recuperación de la mucosa. Como es descrito en el capítulo 4, 66 biopsias de individuos celiacos en remisión y agrupados según las características histológicas de su intestino, fueron elegidos para determinar la evolución temporal virtual de los genes expresados durante la regeneración tisular. Esto nos permitió determinar tanto los genes diferencialmente expresados y con una función en la respuesta inmune como los cambios dinámicos de su patrón de expresión durante el proceso de regeneración. Los resultados de este experimento mostraron que los patrones de expresión de los genes diferencialmente expresados tienen un aumento o una disminución gradual desde el momento en el que comienza el tratamiento con la dieta hasta la completa recuperación del intestino. Por lo tanto, esta observación no refleja como independientes los estadios que clasifican la enfermedad y que son usados para su diagnóstico (Marsh III). Estos resultados indican que a pesar de las diferencias histológicas, los pacientes en el primer y segundo estadio (Marsh I y II) no son diferenciables molecularmente de pacientes en el último estadio (Marsh III). Por ello, proponemos que individuos en el estadio I y II deberían ser considerados como enfermos celiacos siempre que respondan favorablemente al tratamiento con una dieta sin gluten. La clasificación de los genes diferencialmente expresados como parte de la respuesta inmune adaptativa o innata, mostró que ambas respuestas están igualmente representadas en los resultados. Además, estos resultados sugirieron que el interferón gamma es el responsable del mantenimiento crónico de la respuesta inmune adaptativa mediante la activación de la cascada JAK-STAT1 y consecuentemente, de la inducción de la transcripción de quemoquinas y citoquinas que atraen y activan la respuesta adaptativa Th1. Con relación a la respuesta inmune innata, los estudios de expresión génica evidenciaron el reclutamiento de neutrófilos en la lámina propia. Estos resultados fueron corroborados en estudios de inmunohistoquímica en los que un anticuerpo específico para neutrófilos fue usado en biopsias de Marsh III y pacientes celiacos totalmente recuperados (Marsh 0). Estos experimentos indicaron que los individuos celiacos tienen un estado permanente de activación de la respuesta innata en la lámina propia y que es independiente de la presencia o ausencia del gluten en la dieta o del grado de daño tisular. Basándonos en esta observación, llegamos a la hipótesis de que esta infiltración de neutrófilos podría ser el reflejo del efecto causal del gen de la miosina 9B. La miosina 9B podría estar involucrada en la discapacidad de la barrera intestinal y por tanto ser responsable del aumento en la permeabilidad que padecen los individuos celiacos.

La integración de los resultados de los estudios de expresión y de los estudios de ligamiento, junto con el conocimiento de la patogenia de la celiaca, nos llevó a proponer dos genes candidatos funcionales que fueron investigados en estudios de asociación genética (capítulos 5 y 6). El primer gen codifica una enzima llamada proly-endopeptidasa (PREP) que es capaz de fragmentar péptidos del gluten enriquecidos con residuos de prolina. Para la realización

de este estudio, partimos de la hipótesis de que la actividad alterada de esta enzima podría ser responsable de una fragmentación incompleta del gluten, dando lugar a la acumulación de péptidos de gluten capaces de activar la respuesta inmune y por tanto, participar en la pérdida de tolerancia oral al gluten. Para investigar la contribución de PREP como gen causal en la enfermedad celiaca, también llevamos a cabo estudios de secuenciación y de determinación de la actividad de la enzima. Los resultados de los estudios genéticos de asociación y la determinación de la actividad enzimática indicaron que PREP no es un gen causal en la población celiaca holandesa (capítulo 5). El capítulo 6 describe el estudio genético de asociación realizado en el gen candidato funcional y posicional, el factor de transcripción y activador de transcripción 1 (STAT-1), localizado en el pico de ligamiento del cromosoma 2. Este gen codifica uno de los moduladores más importantes de la actividad del interferón gamma. Basándonos en esta información, especulamos que cambios genéticos que diesen lugar a un cambio en la actividad transcripcional de este gen podrían explicar el efecto nocivo de la respuesta inmune que se activa en la celiaca. Sin embargo, no observamos evidencia de una asociación genética significativa en la población celiaca holandesa.

En conclusión, los estudios de expresión génica realizados en este estudio usando biopsias de individuos que padecen la enfermedad celiaca y controles de las mismas, han dado lugar a la identificación de nuevos procesos que intervienen en la patogenia de esta enfermedad. Además, hay que tener en cuenta que la integración de información genética y genómica será clave para desarrollar nuevos métodos de diagnóstico, para identificar individuos con riesgo de desarrollar la enfermedad y nuevas formas de intervención terapéutica en la enfermedad celiaca.

SAMENVATTING

Coeliakie is een genetische aandoening waarbij omgevingsfactoren en meerdere genen bijdragen aan de ontwikkeling van de ziekte. De belangrijkste omgevingsfactor welke een rol speelt bij coeliakie is het eiwit gluten. Dit eiwit is aanwezig in graan, zoals tarwe, gerst en rogge en het zorgt ervoor dat we alledaagse producten zoals brood of pasta kunnen bewerken.

De enige twee tot nu toe bekende genetische factoren zijn genen waarvan er één op chromosoom 6 ligt en codeert voor het HLA-DQ2/DQ8 molecuul en het andere op chromosoom 19 ligt en codeert voor een eiwit genaamd MYO9B.

Klinische symptomen varieren van typische diarree, abnormale absorptie van voedingsstoffen of een groei achterstand tot vermoeidheid, bloedarmoede, gewichtsverlies of helemaal niets. De enige therapie bestaat uit een levenlang glutenvrij dieet.

De ziekte manifesteert zich indien een patient gluten binnenkrijgt en dit zich door het slijmvlies van de dunne darm werkt. Daar aangekomen is gluten in staat om een abnormale immuunrespons te activeren welke zal gaan leiden tot verschillende veranderingen in het slijmvlies: lymfocytose (Marsh I), een abnormale groei van de crypten (Marsh II) en een progressieve afbraak van de darmvlokken (Marsh III).

De kennis over de ontwikkeling van de ziekte is in de laatste jaren flink gegroeid. We weten nu onder andere dat gluten zowel het aangeboren als het verworven immuunsysteem kan activeren. Dit neemt echter niet weg dat diepere kennis van het karakter van de immuunrespons en het mechanisme achter de verandering van de opbouw van het weefsel in de dunne darm nog steeds niet aanwezig zijn.

Het in dit proefschrift beschreven onderzoek voorziet in nieuwe inzichten in de moleculaire mechanismen van zowel het verworven als het aangeboren immuunsysteem en de mechanismen welke verantwoordelijk zijn voor de vlokatrofie.

Het aan het licht brengen van nieuwe aspecten van de patho-physiologie van de ziekte was de belangrijkste impuls in de zoektocht naar kandidaatgenen die iemand vatbaar maken voor coeliakie.

Hoofdstuk 3 beschrijft de transcriptionele veranderingen in de dunne darm van coeliakie patienten met afgebroken darmvlokken (Marsh III) ten opzichte van gezonde controles. We vergeleken expressie profielen van 15 goed gekarakteriseerde Marsh III en zeven controle biopten in een microarray experiment waarbij gekeken werd naar 19.500 genen.

We lieten zien dat een set van 109 genen betrokken is bij lange termijn afbraak van het weefsel. Een groot aantal van deze genen heeft een functie in proliferatie en differentiatie cascades en zijn mogelijk betrokken bij de homeostase binnen de darmvlokken. Dit werd verder ondersteund door de expressiepatronen van Marsh III patienten op een glutenvrij dieet te vergelijken met gluten etende Marsh III patienten. Deze analyse resulteerde in een tweede set van 120 genen welke duiden op verminderde mitotische activiteit in Marsh III patienten na het wegnemen van gluten uit het dieet. Gebaseerd op deze resultaten durfden wij te stellen dat de beide verschillende processen van abnormale groei van de crypten en vlokatrofie moleculair verweven zijn met elkaar. De hypothese is dan ook dat cellen in de crypten niet in staat zijn om hun terminale differentiatie programma af te maken om zo naar de vlokken toe te migreren wat uiteindelijk verantwoordelijk is voor het vormen van de vergrootte crypten

en vervolgens vlokatrofie. Een tweede microarray experiment werd uitgevoerd om specifiek de transcriptionele activiteit van de genen gerelateerd met de immuunrespons tijdens het vernieuwen van het darmslijmvlies te meten.

Zoals beschreven in hoofdstuk 4 hadden wij de mogelijkheid om, met behulp van 66 biopten van coeliakie patienten in verschillende stadia van herstel en van controles, een virtueel tijdsreeks experiment uit te voeren waarmee we zowel verschil in expressie van immuun gerelateerde genen als de dynamiek van hun expressiepatroon tijdens weefselregeneratie kunnen bepalen.

De resultaten van dit experiment gaven aan dat expressiepatronen van genen welke verschil in expressie vertonen, een geleidelijke afname of toename lieten zien vanaf het moment dat de behandeling was gestart tot het moment van volledig herstel. Interessant was dat deze observatie niet strookte met de afzonderlijke histologische stadia waarmee de diagnose gesteld wordt. Deze resultaten wezen erop dat, ondanks morfologische verschillen, Marsh I en Marsh II patienten op moleculair niveau gezien niet verschillen van Marsh III patienten. Zodoende stelden we voor dat deze patienten gezien moeten worden als coeliakie patienten zolang ze goed reageren op een glutenvrij dieet.

De classificatie van genen betrokken bij de aangeboren of verworven immuunrespons lieten een even grote betrokkenheid zien van beide immuunresponsen indien gluten uit het dieet wordt verwijderd.

Onze resultaten gaven indicaties dat interferon gamma de onstekingsreactie kan handhaven door stimulatie van de transcriptie van chemokines en cytokines welke, via de JAK-STAT1 cascade, de Th1 immuunrespons activeren.

Wat betreft de aangeboren immuunrespons lieten expressie studies genen zien die betrokken zijn bij het aantrekken van neutrofielen in de lamina propria van coeliakie patienten. Dit resultaat werd verder ondersteund door middel van immunohistochemie waarbij neutrofielen met behulp van specifieke antistoffen in zowel Marsh III patienten als in patienten in complete remissie (Marsh 0) werden aangetoond.

Deze resultaten impliceerden dat coeliakie patienten constant over een actief aangeboren immuunsysteem beschikken, ongeacht het aan- of afwezig zijn van weefselschade en blootstelling aan gluten.

We speculeerden dan ook dat het verhoogde neutrofielen infiltraat het onderliggend effect van het MYO9B-eiwit onthult. MYO9B kan betrokken zijn bij een verminderde intestinale integriteit en kan daardoor een rol spelen in de verhoogde permeabiliteit van de darmen zoals aanwezig in coeliakie.

Het integreren van gen-expressie data met bekende linkage data, gecombineerd met kennis van het ziektebeeld van de ziekte, leidde ons naar twee functionele kandidaatgenen die vervolgens werden getest in associatiestudies zoals beschreven in hoofdstuk 5 en 6.

Het propyl-endopeptidase (PREP) is een enzym dat in staat is om proline-rijke glutenpeptides af te breken. We stelden voor dat een veranderde PREP activiteit verantwoordelijk zou kunnen zijn voor een inefficiente afbraak van glutenpeptides, wat weer kan leiden tot een opéénstapeling van immuunsysteem stimulerende glutenpeptides en dus een rol zou kunnen spelen in de ontmanteling van de glutentolerantie.

Om de rol van het PREP-gen als primaire factor in coeliakie te onderzoeken, hebben we een sequentie-analyse en een genetische associatiestudie uitgevoerd en hebben we de PREP-enzymactiviteit bepaald in biopten van coeliakie patienten en controles. Onze resultaten van de genetische associatiestudie en de activiteitsmetingen lieten zien dat PREP niet het ziekteveroorzakende gen is in de nederlandse coeliakie populatie, zoals beschreven in hoofdstuk 5.

Hoofdstuk 6 beschrijft de associatiestudie welke is uitgevoerd op de functionele kandidaat signal transducer and activator or transcription 1 (STAT-1), welke ook nog gelokaliseerd is onder de linkage piek op de lange arm van chromosoom 2.

Dit eiwit wordt gereguleerd door interferon gamma. We speculeerden dat elke genetische variatie die de transcriptionele activiteit van dit eiwit zou veranderen, verantwoordelijk zou kunnen zijn voor de schadelijke immuunrespons zoals we zien in coeliakie.

Echter, er werd opnieuw geen bewijs gevonden voor positieve associatie binnen de nederlandse populatie.

De gen-expressie studies die in dit project zijn uitgevoerd op biopten van coeliakie patienten en controles hebben geleid tot de identificatie van nieuwe cascades die betrokken zijn bij het ziektebeeld van coeliakie.

Deze verkregen kennis zal zeker bijdragen aan het onderzoek naar genen betrokken bij het ontstaan van coeliakie.

Bovendien zal de integratie van genetische en genomische informatie de sleutel zijn tot de ontwikkeling van diagnostische handvatten die bij zullen dragen aan de identificatie van personen met een verhoogd risico op coeliakie en tot nieuwe therapeutische mogelijkheden leiden.

ACKNOWLEDGEMENTS/ AGRADECIMIENTOS

Cisca, thank you very much for giving me the opportunity to taste science. I have worked in your lab with big fun, both in the last years with all the successful results, and during those previous, hard-working times in which you cheered us up with your enthusiasm, energy and good ideas. This has all been a terrific experience. I will carry a lot of good and wise advice from you into my postdoc time.

Chris Mulder, Jos Meijer and Gerrit Meijer, thank you very much for all your enthusiasm in the coeliac project at all times and for your dedication to it in between your daily clinical practice. This project could never have been possible without your interest and confidence in our research. Bart Crusius, thank you very much for all your HLA info, and Salvador Peña, it was always really cheering to hear from you how good the wines produced in Valladolid are and to speak a bit of Spanish with you.

Of course, my thanks to everybody who has contributed to the coeliac disease project in Leiden, Wageningen, the coeliac patients, and specially to Martine, Alfons, Karen, Martin, Sasha, Erica, Lude, Alienke, Dario, Jan, Mariana and Ruben. I really enjoyed working with you all in this project, the lab work, writing papers, brainstorming, deadlines, presentations and congresses. But besides the work there were your funny stories, your humor, your personal experiences, the shared ice-creams, the gluten-free diet experiment, dinners… they all tightened our bonds to the coeliac group. I will miss you all.

Jackie, thank you very much for all the editing work. Even when we all wanted our papers edited at the same time and ready yesterday, of course, you never got upset. It was always fun to comment with you about the peculiarities of the Dutch while cycling back home. You also made possible a really fantastic barbecue last summer and, believe it or not, your botanic challenge is still alive in the flexroom!

Bart, Prim, Patricia, Harm, Bart J, Dineke, Steven, Bobby, Linda, Birgit, Dalila, Judith, Carline, Tineke, Leontine, Gustavo, Jonathan, Behrooz, Harry, Ynte, Albertine and Flip, thank you all for all your help with my project and future plans at the work discussions and the lab retreats, and for all the enjoyable *sjoelen* and ping-pong competitions, barbecues, celebrations for birthdays, marriages, births, Sinterklaas, good-bye parties, published papers, *promoties*, awarded grants, blunder cakes… yes, there were a lot of calories ingested!

Por supuesto, a la delegación granadina en Utrecht: Blanca y Elena. Gracias por traerme un poco de sol español a esta Holanda y confirmarme que no es normal lo de desayunar, comer, y hasta cenar queso todos los días, sacar la agenda para quedar con un amigo o ir en bicicleta cuando diluvia o con media casa a cuestas... Garrote y Eduardo: me dio mucha alegria coincidir en el congreso de la celiaca en París y poder charlar con vosotros. Garrote, gracias por hacerme un hueco el verano pasado durante tu visita en Amsterdam y a Eduardo y Alberto por tan buena sobremesa en Oporto. María Luisa, ha sido toda una suerte conocerte, no solo me has dado consejos profesionales en las comisiones de evaluación, sino que tu experiencia y tu buen hacer en Holanda son sin duda todo un ejemplo. A todos vosotros, gracias por haberme hecho recordar todas esas pequeñas cosas y maneras que se echan en falta o incluso se olvidan cuando estas lejos de tu tierra y que llena tanto recordar.

The Bilthuis friends, Marta, Bartje, Marco, Federico, Pia, Stefano, Sofia…lots of fun, experiences, conversations, international dinners, trips, parties. This was definitely a very

intense and unforgettable time of our lives. Luckily we keep in touch despite the distances and time.

Helena, I think it has been our great luck to be friends and to share this experience at the same time. We definitely supported each other a lot, even until the last minute of the whole adventure! Thank you, too, for our long therapeutical talks. Arantza, gracias por todas las visitas, los lunch de los sábados y sobre todo por los buenos consejos sobre tu experiencia en Holanda.

Isa, Manuel, Raquel, Vicky, Rafa, Virginia, Alfredo, Miriam, Jeremy, Olga Gallo, Olga Fer, Michi, Antonio, Rafa, Noe, Victoria. Recordar nuestros tiempos en la biblioteca, los sábados en la calle Paraíso, las noches viejas, el colegio, los días en el instituto siempre me traen un buen recuerdo y una sonrisa. Dicen que la distancia es el olvido, nuestra experiencia ha demostrado definitivamente que esos dichos de nuestras abuelas no siempre funcionan. Tenemos una celebración pendiente.

Papá, mamá, Bea y Rodrigo, como no, esta tesis está dedicada a vosotros, a las personas que más quiero, a vuestra ayuda incondicional, a mi familia.

Eric, you definitely deserve my biggest thanks. You have always been there during these past four years. Thank you for all your understanding and your good advice; it has no price. I will miss not to cycle with you everyday and enjoy together the beautiful Dutch nature.

Curriculum Vitae

Maria Begoña Diosdado Calvo was born 30th January year 1976 in Valladolid, Spain. She started medical school at the University of Valladolid in 1994. During the summer of the third year she had the opportunity to collaborate in a project that aimed to detect mutations in hereditary colon cancer at the Department of Biochemistry, in the Medical School, University of Nuevo León, Mexico. During the summers of her fourth and fifth years, she did a traineeship at the Department of Biochemistry and Physiology, University of Valladolid. The aim of this study was to identify mutations in the hereditary breast cancer genes, BRCA1 and BRCA2. During the last year of her study she was awarded a Leonardo grant from the European Union to work at the Department of Medical Genetics, University Medical Center Utrecht, in the Netherlands. The aim of this work was to identify the genetic variation in the calpain-10 gene in the type 2 diabetes mellitus population of Breda, under the supervision of J. van Tilburg and Dr. C. Wijmenga. This led to a PhD place in the Complex Genetics Section, under the supervision of Prof. C.Wijmenga, and the results described in this thesis . Since 1st February 2006 she has held a postdoctoral fellowship in the Tumor Profiling Group led by Prof. Gerrit Meijer, in the Department of Pathology, Free University Amsterdam.

List of Publications

This thesis

Diosdado B, van Bakel H, Strengman E, Franke L, van Oort E, Mulder CJ, Wijmenga C, Wapenaar MC. A genomics view on celiac disease points to enhanced neutrophil recruitment and barrier impairment. (submitted).

Diosdado B, Monsuur AJ, Mearin ML, Mulder CJJ, Wijmenga C. The down-stream modulator of interferon-gamma, STAT1 is not genetically associated to the Dutch coeliac disease population. (submitted).

Diosdado B and 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 Sep;5(5):681-700.

Diosdado B, van Oort E, Wijmenga C. "Coelionomics": towards understanding the molecular pathology of coeliac disease. Clin Chem Lab Med. 2005;43(7):685–695.

Diosdado B, Stepniak DT, Monsuur AJ, Franke L, Wapenaar MC, Mearin ML, Koning F, Wijmenga C. No genetic association of the human prolyl endopeptidase gene found in the Dutch celiac disease population. Am J Physiol Gastrointest Liver Physiol. 2005 Sep;289(3): G495-500.

Diosdado B, Wapenaar MC, Franke L, Duran KJ, Goerres MJ, Hadithi M, Crusius JBA, Meijer JWR, Duggan DJ, Mulder CJJ, Holstege FCP, Wijmenga C. A microarray screen for novel candidate genes in celiac disease pathogenesis. Gut. 2004 Jul;53(7):944-51.

Others

Monsuur AJ, Stepniak DT, Diosdado B, Wapenaar MC, Mearin ML, Koning F, Wijmenga C. The PGPEP1 gene and its product pyroglutamyl peptidase I are not implicated in the pathogenesis of celiac disease. (submitted).

Zhernakova A, Alizadeh BZ, Eerlig P, Hanifi-Moghaddam P, Diosdado B, Wijmenga C, Roep BO, Koeleman BPC. Genetic variants of RANTES are associated with serum RANTES level and protection to T1D. (submitted).

Monsuur AJ, de Bakker PIW, Alizadeh B, Bevova M, Franke L, t' Slot R, van Belzen MJ, Diosdado B, Daly MJ, Mulder CJJ, Mearin MJ, Meijer JWR, Meijer GA, van Oort E, Wapenaar M, Koeleman BPC, Wijmenga C. Myosin IXb variant increases the risk of celiac disease. Nat Genet. 2005 Dec;37(12):1341-4.

Franke L, van Bakel H, Diosdado B, van Belzen MJ, Wapenaar M, Wijmenga C. TEAM: a tool for the integration of expression, and linkage and association maps. Eur J Hum Genet. 2004 Aug;12(8):633-8.

COLOR FIGURES

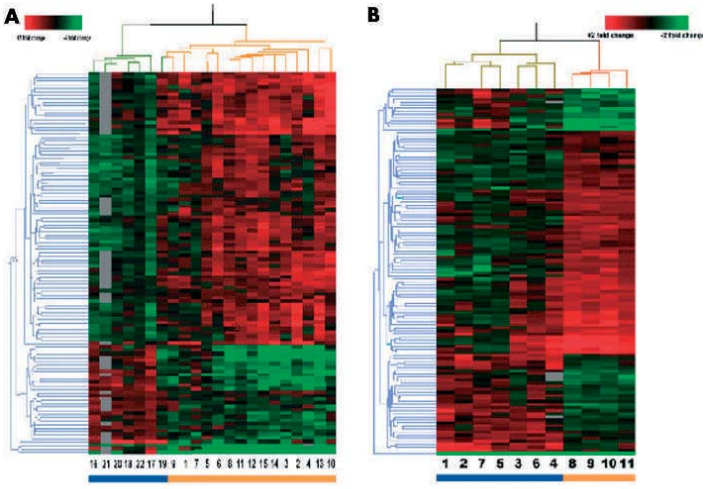
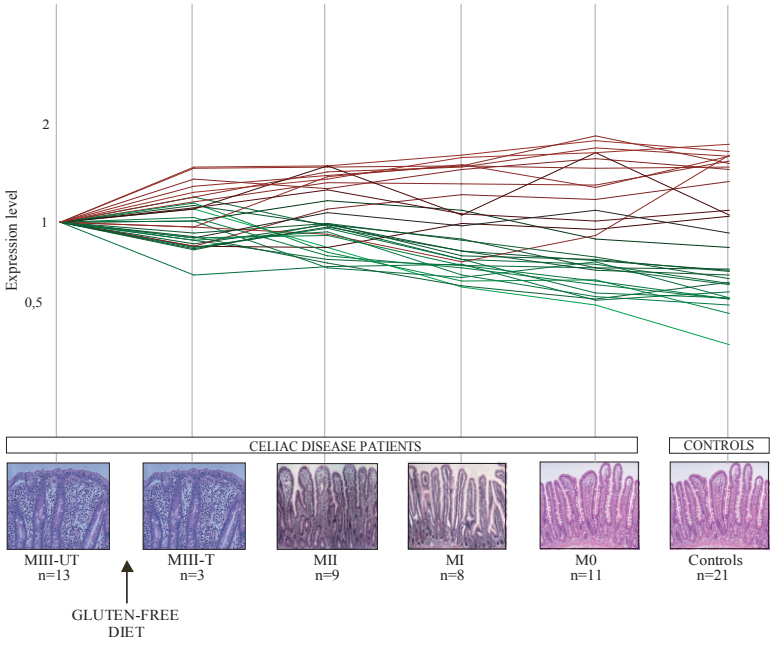
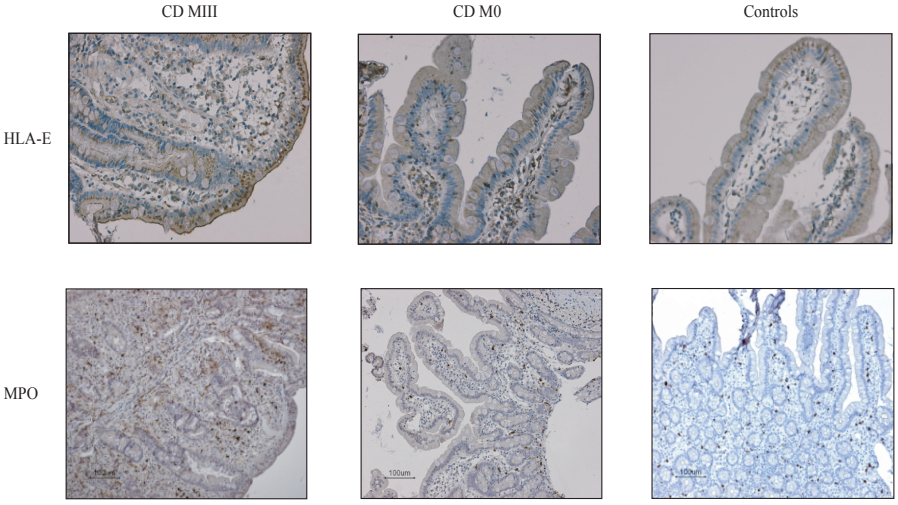


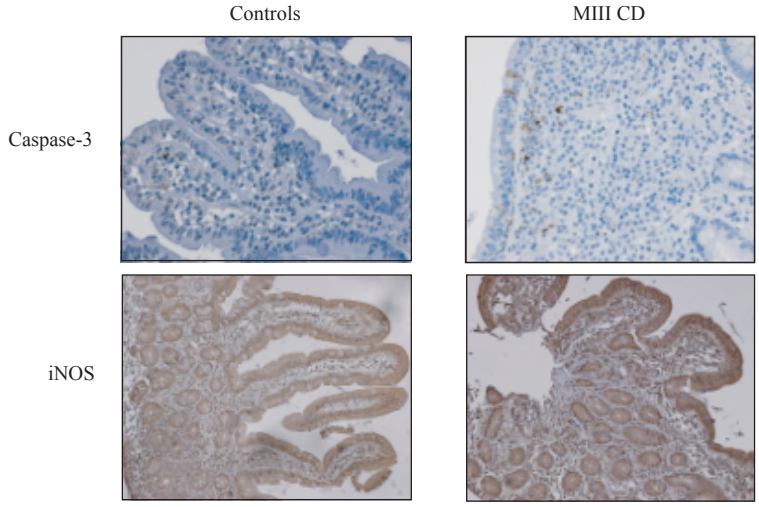
Figure 1 (A) Hierarchical clustering dendrogram of duodenal genes from Marsh III (MIII) and Marsh 0 (M0) biopsies. Clustering of 109 genes across 22 samples clusters the seven control samples (blue bar) separately from the 15 coeliac disease patients (orange bar). Each column represents a coeliac disease (MIII) or a control (M0) sample and each row represents an individual gene. For each gene, a green signal represents underexpression, black signals denote similarly expressed genes, a red signal represents overexpressed genes, and grey signals denote missing data. Information on these genes can be found in table 3 and on the website http://humgen.med.uu.nl/publications/CD/Microarrays/Diosdado_1.html. (B) Hierarchical clustering dendrogram of duodenal genes from MIII biopsies with or without exposures to gluten. Clustering of 120 genes across 11 samples clusters the seven samples of coeliac disease patients following a gluten containing diet (blue bar) separately from the four coeliac disease patients on a gluten free diet (orange bar). Each column represents a sample and each row represents an individual gene. For each gene, a green signal represents underexpression, black signals denote similarly expressed genes, a red signal represents overexpressed genes, and grey signals denote missing data. Information on these genes can be found in table 3 and on the website http://humgen.med.uu.nl/publications/CD/Microarrays/Diosdado_1.html.



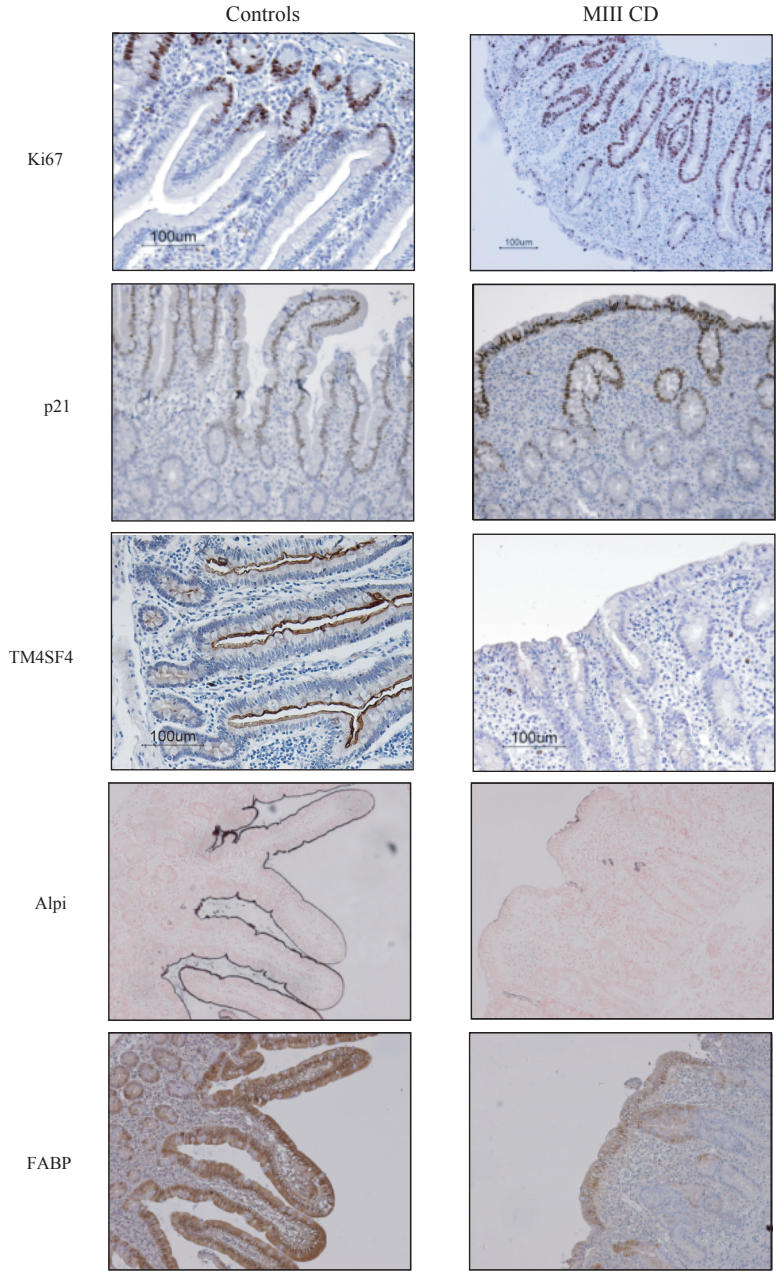
Chaper 4, figure 1. Profiles of differentially expressed immune-related genes during mucosal recovery. The pattern of expression of 27 immune-related genes selected by MAANOVA analysis showed a gradual change as opposed to the discrete histological transitions during remission. The upper panel shows the relative expression of genes that was either enhanced (red) or decreased (green) during remission. Expression was normalized to the values measured in untreated Marsh III patients. The lower panels show the corresponding histology and Marsh classification of the biopsy samples taken before treatment and during remission on a gluten-free diet. Histopathological features include villous atrophy (MIII), crypt hyperplasia (MIII-MII), and lymphocytosis (MIII-MI). Patients in complete remission (M0) are comparable to controls. MIII-UT: Marsh III untreated; MIII-T: Marsh III treated; MII: Marsh II; MI: Marsh I; M0: Marsh 0. Indicated is the number of independent individuals included in each group.



Chapter 4, figure 2. Immunohistochemical localization of HLA-E and myeloperoxidase (MPO) in the duodenal mucosa of coeliac patients and controls. Shown is the expression of the non-classical MHC class I molecule HLA-E and the neutrophil-specific marker MPO in patients (Marsh IIII and Marsh 0) and healthy controls. A) HLA-E was strongly expressed in the Marsh III epithelial layer but absent in normalized patients (Marsh 0) and controls. No differences in staining were found at the lamina propria. B) Increased number of MPO-positive cells in the lamina propria of Marsh III and Marsh 0 patients compared to controls. This indicated a higher number of neutrophils in the coeliac mucosa regardless of the histological stage or the gluten-free diet followed.



General discussion, figure 2. Immunohistochemistry studies on pathways postulated to drive villous atrophy in coeliac disease. The presence of apoptosis in the villi of CD patients and controls was assessed by anti-caspase3 antibody, which detects the activated form of the apoptotic-related caspase-3. No differences in inmmunoreactivity were detected between patients and controls. Inducible nitric oxide synthase (iNOS) staining in the mucosa of CD patients showed increased inmmunoreactivity in the lamina propria of the MIII CD patients compared to controls. No differences were found at the epithelial layer.



General discussion, figure 3. Immunohistochemistry studies on proliferation and differentiation pathways in coeliac disease. The proliferation state of the cells of the crypts in CD patients and controls was evaluated with the ki-67 marker, which showed increased cellularity of the crypts in patients compared to controls. To further examine the differentiation state of the cells of the villi, the marker p21, which shows cell arrest, was tested but showed no differences between MIII CD patients and controls. However, the markers for terminal differentiation tetraspanin family member 4 (TM4SF4) and alkaline phosphatase (Alpi) did show decreased and patchy staining in biopsies from MIII patients compared to controls, although the fatty acid-binding protein (FABP) showed no differences in inmmunoreactivity.