

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

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.

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

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 improvement⁸ 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 compliant¹⁰ 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 Darray suite¹¹ 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 ¹⁰log 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, Emssembl 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^{-ΔΔCt} 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.

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

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.

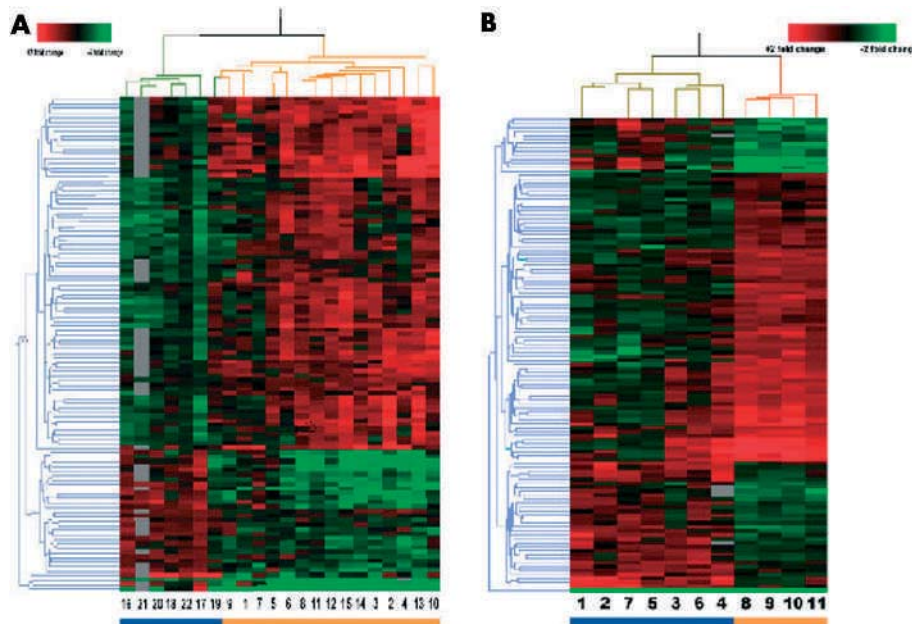


Figure 1 (A) Hierarchical clustering dendrogram of duodenal genes from Marsh III (MIII) and Marsh 0 (M0) biopsies. Clustering of 109 genes across 22 samples clusters the seven control samples (blue bar) separately from the 15 coeliac disease patients (orange bar). Each column represents a coeliac disease (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 (*LCPI*) gene. Genes involved in complementary functions, such as complement component 8 β polypeptide (*C8B*) and the H factor 1 (*HFI*), 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 adjuvant 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 $\gamma\delta$ 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 $\gamma\delta$ T cells.³ Hence these results may reflect the increased number of $\gamma\delta$ 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*), synaptotodin (*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 $\gamma\delta$ 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				
CSR1	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				
HSPA6	Pro-neuregulin-3 precursor		10q23.1	0.69
SARM	Heat shock 70 kDa protein 6	3310	1q23.3	0.76
	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	Hpall 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 meningoseticum* 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 Dariusz Stepniak for helpful discussions, Daniel Chan for helping with the illustration, Justin Franssen 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

- Vader W, Kooy Y, Van Veelen P, *et al*. The gluten response in children with coeliac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* 2002;**122**:1729-37.
- Sollid LM. Molecular basis of coeliac disease. *Annu Rev Immunol* 2000;**18**:53-81.
- 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.
- Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* 2002;**2**:647-55.
- Schuppan D. Current concepts of coeliac disease pathogenesis. *Gastroenterology* 2000;**119**:234-42.
- Lockhart DJ, Winzler EA. Genomics, gene expression and DNA arrays. *Nature* 2000;**405**:827-36.
- Mulder C, Rostami K, Marsh MN. When is a coeliac a coeliac? *Gut* 1998;**42**:594.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Nilsen EM, Jahnsen FL, Lundin KE, *et al*. Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with coeliac disease. *Gastroenterology* 1998;**115**:551-63.
- 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.
- Yakota T, Ehlin-Henriksson B, Hansson GK. Scavenger receptors mediate adhesion of activated B lymphocytes. *Exp Cell Res* 1998;**239**:16-22.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Shan L, Molberg O, Parrot I, *et al*. Structural basis for gluten intolerance in coeliac sprue. *Science* 2002;**297**:2275-9.
- Wice BM, Gordon JL. 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.
- Battle 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.
- 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.