Transcriptome Profiling and Genetic Analysis to Identify Susceptibility Genes for Crohn’s Disease

Debby LAUKENS

Gent, June 2006

Thesis submitted in partial fulfilment of the requirements for the degree of

DOCTOR IN SCIENCES: BIOTECHNOLOGY

Promoters: Prof. Erik Remaut and Dr. Pieter Rottiers
Examination commission

Chairwoman

Prof. Marcelle Holsters

Reading committee

Prof. Erik Remaut
Dr. Pieter Rottiers
Prof. Martine De Vos
Prof. Dirk Elewaut
Prof. Jurgen Del-Favero

Other members

Prof. Lothar Steidler
Prof. Filip De Keyser
Dr. Pieter Demetter

1 Department of Plant Systems Biology, Ghent University and Flanders Interuniversity Institute for Biotechnology (VIB), Gent, Belgium
2 Department for Molecular Biomedical Research, Ghent University, VIB, Gent, Belgium
3 Department of Gastroenterology, Ghent University Hospital, Gent, Belgium
4 Department of Rheumatology, Ghent University Hospital, Gent, Belgium
5 Department of Molecular Genetics, University of Antwerp, VIB, Antwerpen, Belgium
6 Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland
7 Department of Pathology, Erasmus University Hospital, Brussel, Belgium
# Scope of the project

### Section I

**Chapter 1**

Introduction to the molecular genetics of Crohn’s disease

**Chapter 2**

A functional promoter polymorphism in *CARD15*
*Manuscript in preparation*

**Chapter 3**

Reduced metallothionein expression in colonic Crohn’s disease: evidence for *MTF1* as a new disease-modifying gene
*Submitted to Gastroenterology*

**Chapter 4**

Human metallothionein expression under normal and pathological conditions: mechanisms of gene-regulation
*Review in preparation*

### Section II

**Chapter 5**

Spondyloarthropathy as a model for early Crohn’s disease

**Chapter 6**

*CARD15* mutations in patients with spondyloarthropathy are linked with disease progression and evolution to Crohn’s disease

**Chapter 7**

Altered gut transcriptome in Spondyloarthropathy
*Ann Rheum Dis.* 2006 Feb

**Future prospects**

**Summary and discussion**

**Samenvatting en discussie**

**Curriculum vitae**

**Dankwoord**
Schematic outline of the project. A central theme in this study was the use of gene expression analysis in the search for susceptibility genes for Crohn’s disease (CD), and for the study of CD-related gut inflammation observed in spondyloarthropathy (SpA). Crohn’s disease is a complex genetic trait where both genes and environmental factors play a role. In Chapter 1, the current status on genetic susceptibility in CD is reviewed. A novel functional promoter polymorphism in the recently identified CD-susceptibility gene, CARD15, is illustrated in Chapter 2. We describe a method to identify new susceptibility genes by the use of transcriptome profiling in Chapter 3, while two new candidate genes are characterized and described in Chapters 3 and 4 respectively. The occurrence of silent or subclinical gut inflammation in SpA patients and its relation to CD is reviewed in Chapter 5. In addition, the genetic contribution of CARD15 in SpA is studied in Chapter 6. In Chapter 7, we compared global gene expression in intestinal biopsies of SpA and CD patients, and provided additional evidence that SpA is a good model to study early CD. Finally, the influence of genetic risk factors on gene expression is described in more detail in the Future prospects, together with an allusion on the importance of epigenetic mechanisms in gene expression studies. In this last section, a new approach for the identification of susceptibility genes, based on the transmission of gene expression within families, is suggested.
Section I
The objective of this project was to use gene expression analysis for the study of genetic susceptibility in Crohn’s disease (CD). Crohn's disease is a painful and incapacitating chronic inflammatory disease of the intestine. The symptoms of CD include chronic diarrhea, abdominal pain, loss of appetite, and weight loss. Approximately 1-2 in 1,000 people are affected and the incidence is still increasing. A life-long treatment is required, and the quality of life is severely influenced. Currently, there is no cure and patients suffer from a high rate of recurrence.
The cause of ongoing inflammation in the intestine of patients with CD results from a complex interaction of susceptibility genes and environmental factors. Chapter 1 provides an overview of the current status of genetic linkage data, association studies and genetic risk factors for CD. In 2001, the first susceptibility gene for CD was cloned. This gene, called CARD15, is an intracellular receptor for bacterial cell-wall components, and can activate NFκB to induce an inflammatory response. Three single nucleotide polymorphisms (SNPs) within CARD15 have been independently associated with CD, and are present in approximately 40% of CD patients. The identification of functional consequences of these SNPs is the subject of much ongoing research. Because CARD15 expression is up-regulated during inflammation, we screened its promoter for polymorphisms. In Chapter 2, we describe a novel SNP in the 5' untranslated region (5'UTR) of CARD15, located within a DNA stretch that contains a potential binding site for the E2F transcription factor. Interestingly, this SNP influences the induction of a reporter gene controlled by a part of the CARD15 promoter in response to tumor necrosis factor alpha, which is a crucial pro-inflammatory cytokine. Furthermore, this polymorphism is linked to the known CD-associated mutations. In order to characterize this link, we are currently mapping the haplotype structure surrounding CARD15.
Unfortunately, the identification of new susceptibility genes for CD by classical linkage studies is complicated by genetic heterogeneity and incomplete penetrance of the disease. Therefore, in Chapter 3, we suggest an alternative approach for the identification of new candidate genes by the use of transcriptome analysis. This initiative was based on the idea that much of the natural variation in gene expression is genetically transmitted and probably accounts for susceptibility to common diseases such as CD. We studied genome-wide gene expression in biopsies from unaffected colon areas of CD patients and healthy controls, and focused on those genes that are differentially expressed and located near a known locus for CD. A cluster of highly related genes, encoding metallothioneins (MTs), was studied in more detail. These genes are located approximately 6 centimorgan q-telomeric from the locus containing CARD15. We provide evidence for MT expression as a genetic risk factor in CD, and describe a new disease-
Scope of the project

modifying polymorphism in the main transcriptional regulator for MTs, called the MRE-binding transcription factor 1 (MTF1). In addition, a review of the involvement of MT in several diseases is provided in Chapter 4, together with current knowledge of the regulation of human MT expression.

Section II
The most common extra-intestinal manifestation of CD is inflammation of the joints and the spine. This causes pain, swelling and stiffness of the joints and the back. Spondyloarthropathy (SpA) is a group of rheumatic disorders of the spine. Interestingly, a close relationship exists between CD and SpA clinically as well as immunologically, which is discussed in Chapter 5. Not only is CD prevalent among patients with SpA, silent or subclinical gut inflammation has been described in up to two-thirds of patients with SpA. When CARD15 was mapped as the first CD-associated gene, we tested the presence of the CD-associated CARD15 mutations in a cohort of SpA patients, as described in Chapter 6. We provided a first genetic link between CD and SpA: the presence of CD-associated CARD15 polymorphisms was associated with SpA patients with subclinical chronic gut inflammation. Lastly, in Chapter 7, we reported a gene expression survey of intestinal biopsies of SpA and CD patients, and showed that an altered gut transcriptome was found in SpA patients with subclinical chronic gut inflammation. The alterations were comparable to those seen in CD, confirming initial clinical observations and suggesting that SpA serves as a good model for early CD.

To conclude, the mounting significance of studying gene expression in genetic susceptibility studies is outlined in the Future prospects. We describe a new approach to identify risk genes based on the transmission of gene expression signatures in families.
INTRODUCTION TO THE MOLECULAR GENETICS OF CROHN’S DISEASE

Debby Laukens

Department of Molecular Biomedical Research, Ghent University and Flanders Interuniversity Institute for Biotechnology (VIB), Gent, Belgium

INTRODUCTION

Inflammatory bowel disease (IBD) is a term that refers to both Crohn’s disease (CD) and ulcerative colitis (UC). Generally, the diagnosis of CD or UC is made based on the location of inflammation and depth and kind of lesions after endoscopic (Figure 1), radiologic and histologic examination. Although CD most commonly affects the terminal part of the small intestine (ileum) and the proximal part of the large intestine (colon), it may involve any section of the gastrointestinal tract. All layers of the intestine may be involved, and normal healthy bowel can exist between patches of diseased bowel. Granulomas are commonly found. In UC on the other hand, inflammation is limited to the colon. Only the superficial layers (mucosa) of the colon are affected in a symmetric and continuous distribution, starting at the level of the rectum. On the other hand, there may be a spectrum of illnesses within each disorder, suggesting distinct underlying pathogenic mechanisms. In 10 percent of the cases where only the colon is affected, it is difficult to establish a definite diagnosis, hence these are classified as indeterminate colitis (Price 1978). However, most patients with indeterminate colitis evolve to a definite diagnosis of UC or CD during follow-up. Whether UC and CD are fundamentally different diseases, or part of one concept, remains unclear and holds conceptual and practical implications. Good classification practice is essential because CD and UC differ in response to therapy and involvement of genetic and environmental risk factors (Vermeire et al. 2000).

Both CD and UC have a prevalence of 1-2 cases per 1,000 individuals in western countries. In Brussels, the mean annual incidence for CD has been estimated in 1992-1993 at 4.1 per 100,000 inhabitants/year (Van Gossum et al. 1996). The emergence of IBD in the developed nations of
Western Europe and North America during the past century seems indisputable (Farrokhyar et al. 2001). Part of the increase in incidence reflects developments in endoscopic and other diagnostic techniques. Nevertheless, most epidemiologists accept that there was a considerable increase in CD incidence during the last century in developed countries. Western lifestyle, diet (Gibson and Shepherd 2005), exposure to infection (Hugot et al. 2003), and domestic hygiene (Wells and Blennerhassett 2005) are most probably involved in this process. Crohn's disease may occur among people of all ages, but it is primarily a disease of adolescents and young adults, affecting mainly those between the age of 15 and 35.

Currently, the long-term treatment of IBD is multifaceted, depending on the type of disease, and sites involved (Podolsky 2002). Most clinicians use a stepped approach to therapy, the so-called therapeutic pyramid, in which more potent agents are added to the regimen if less active drugs fail to achieve an adequate response. The primary treatment of symptoms with antidiarrheal agents itself is important. Aminosalicylates, corticosteroids and other immunomodulators are aimed at controlling inflammation. The use of antibiotics and probiotics is being investigated, especially for CD. Beneficial health effects, which might result from rebalancing of the gut flora, were observed after such treatment (Sartor 2003). Towards the top of the pyramid are the biological agents. The first biological therapy for CD, based on anti-tumor necrosis factor alpha (anti-TNF) has been approved by the Food and Drug Administration in 1998, and has turned out to be very effective (Targan et al. 1997). Up to 60% of CD patients need an operation within 10 years of disease onset, as a result of fistulizing, perforating, strictureing and/or obstructing complications.

The exact cause of IBD remains uncertain. It is clear that IBD arises from the interplay of environmental factors in genetically defined individuals. Today, the most accepted hypothesis about pathogenesis of IBD is that disease results from an abnormal host immune response to
Introduction to the molecular genetics of Crohn’s disease

bacteria that are normally found in the intestine (Sartor 2003). It is hypothesized that in normal conditions, exposure to commensal bacteria modulates the inflammatory immune response of the gut to the numerous bacteria and food antigens to which it is constantly exposed (Saxelin et al. 2005). In IBD, however, exposure to microflora triggers an inflammatory response by the cells lining the mucosa, leading to a chronic, destructive immune response, ultimately causing ulcerations and bowel injury. Indeed, an immune response to intestinal bacteria seems to be crucial in the pathogenesis of IBD (Sartor 2003): 1) genetically engineered rodents, such as IL10 deficient mice, develop colitis when exposed to commensal bacteria but remain disease free when raised in a sterile environment (Shi and Walker 2004); 2) the negative influence of faecal stream in the pathogenesis of recurrent CD lesions after curative resection of the distal ileum (Rutgeerts et al. 1991); 3) the efficacy of probiotics and antibiotics in treatment and prevention of IBD; 4) loss of tolerance to commensal bacteria (Jump and Levine 2004); 5) the increased number of surface-adherent and intracellular bacteria in the colon and ileum of IBD patients (Swidsinski et al. 2002; Darfeuille-Michaud et al. 2004); 6) the association of microbial receptor gene variants with IBD (Hugot et al. 2001; Klein et al. 2002; Franchimont et al. 2004); and 7) microbial DNA is found in granulomas, the most specific histological lesion found in CD (Ryan et al. 2004). It is, however, unclear whether the immune system is activated as a result of an intrinsic defect (either a constitutive activation or the failure of down-regulatory immune mechanisms) and/or because of continued stimulation resulting from a defective epithelial mucosal barrier.

Many researchers have invested in finding causative agents and risk factors for IBD. One of the leading infectious candidates is *Mycobacterium avium paratuberculosis* (Autschbach et al. 2005), but also *Listeria monocytogenes* and *Helicobacter hepaticus* have been linked to IBD. However, so far, attempts at using specific anti-mycobacterial chemotherapy have been unsuccessful (Goodgame et al. 2001). In any case, the complex phenotypes of IBD cannot be fully explained either by environmental factors, or by a single gene. Increasing evidence suggest a multigenic nature in combination with the above-mentioned aspects.

Crohn’s disease and UC share the same ethnic predisposition, and mixed families exist in which some members are affected with CD and others with UC. However, genetic data of susceptibility genes tend to differ between CD and UC. Here, we review the progress in genetic analyses and genetic risk factors for CD.
Chapter 1

GENETIC EPIDEMIOLOGY OF CROHN’S DISEASE

The strong influence of genetic determinants in CD has been shown by familial clustering (Peeters et al. 1996), and the high concordance rate, about 20-44%, in monozygotic twins (Tysk et al. 1988; Hugot et al. 1996; Thompson et al. 1996; Halfvarson et al. 2003). First-degree relatives of affected individuals show a 20 to 50-fold higher risk for developing CD (Monsen et al. 1991). The sibling relative risk for CD is 36.5, which is higher than the risk reported for other complex diseases such as diabetes type I and schizophrenia (Hugot et al. 1996; Peeters et al. 1996; Satsangi et al. 1996a). Moreover, siblings are frequently affected at similar ages, and concordance rates reach 80% for disease site, behavior and presence of extra-intestinal manifestations (Halfvarson et al. 2003). Ethnic differences in disease frequency have been shown. For instance, the prevalence among Ashkenazi Jews is much higher than among Sephardic Jews, even though they share a similar living environment in the same community (Fidder et al. 2003). An increased mother to child transmission has been observed (Akolkar et al. 1997), suggesting that one of the genetic risk factor is an imprinted gene, which is in agreement with the linkage of CD to the X chromosome shown in several studies (Vermeire et al. 2001; van Heel et al. 2003). All these observations are considered to be evidence for a strong genetic predisposition to the etiology of CD.

The incomplete concordance rate in monozygotic twins suggests that genetic factors are not enough to develop CD. Furthermore, the incidence of CD in immigrant populations is often different from the incidence of the population of origin, but also slightly different from the incidence in the new country (Odes et al. 1989). The strongest environmental risk factor for IBD is tobacco smoking. Remarkably, the risk in UC and CD is different. Smoking is harmful in both the disease progression and in the onset of CD (Cosnes et al. 2001), while it has a positive influence on the symptoms in UC (Beaugerie et al. 2001). In fact, many people who stop smoking develop UC. Risk factors influencing the disease progression include CD affecting the small intestine, age of onset above 40 years and oral contraceptive use (Logan 1999). Recently, environmental factors that might be etiologically related to CD were determined in a study based on interviewing members of affected and control families (Van Kruiningen et al. 2005). Besides smoking and appendicitis, dietary factors represented potential risk factors for CD. These dietary components included drinking of well water, less frequent consumption of oats, rye, and bran and more frequent eating of unpasteurized cheeses. Childhood infections, birth by caesarean section and bottle feeding have all been suggested as possible risk factors for the development for CD (Timmer 2003).
In general, two complementary approaches are used to identify determinants underlying genetically complex traits such as CD: candidate gene association studies and genetic linkage. In the candidate gene approach, functionally interesting genes are screened for polymorphisms in a limited number of patients, followed by comparing frequencies in large disease and control populations. In genetic linkage, candidate regions associated with disease are identified, scanning the entire chromosome for microsatellite markers. These so-called disease loci are usually very broad and can rarely be narrowed down to less than a few megabases. If the candidate region contains an amenable number of genes, a candidate gene is chosen by virtue of a known property, e.g. function, tissue-specific expression or structural motif, which is referred to as positional cloning. However, the mapping of CD loci is hampered by variability in phenotype, genetic heterogeneity across populations, uncontrolled environmental influences and limited statistical power. For instance, in CD, each genome scan results in a minimum of 3 loci, reflecting that more than one gene is involved in pathogenesis. The number of genes involved in the etiology of CD is unknown. Furthermore, the question remains whether common variants of a limited number of genes, each adding up small effects, account for such a common disease, or whether rare variants with a dominant effect underlie genetic susceptibility.

Although not all CD patients have a positive family history, the question arises as to the precise relationship between sporadic and familial cases of CD. Will genetic variants identified in familial cases also be pathogenic in sporadic CD patients? In a recent study addressing this question, no difference in frequency of three CD-associated variants within the \textit{CARD15} gene (see below) was found between familial and non-familial cases (Vermeire \textit{et al.} 2002). This suggested that these variants are partly causative for idiopathic CD patients, and are also transmitted within CD families (Esters \textit{et al.} 2004).

**GENOME-WIDE LINKAGE STUDIES IN IBD**

At present, 13 genome-wide linkage studies have been performed in IBD (Hugot \textit{et al.} 1996; Satsangi \textit{et al.} 1996b; Cho \textit{et al.} 1998; Hampe \textit{et al.} 1999; Ma \textit{et al.} 1999; Duerr \textit{et al.} 2000; Rioux \textit{et al.} 2000; Cavanaugh 2001; Paavola-Sakki \textit{et al.} 2003; Shaw \textit{et al.} 2003; van Heel \textit{et al.} 2003; Barmada \textit{et al.} 2004; Vermeire \textit{et al.} 2004). More than 20 loci have been identified, and many of them have been independently replicated (\textit{IBD1-9}, 3q, 4q, 6q, 7q, 8q, 10p, 22q, Xp). Table 1 covers all original articles on linkage analysis performed on IBD or mixed families from 1996 until now. Data on UC only were not included. These analyses uncover great ethnic differences as to loci and degree of linkage. Identification of the precise genes within these loci
## Table 1 Susceptibility loci for Crohn’s disease

<table>
<thead>
<tr>
<th>IBD Locus</th>
<th>Chromosomal position</th>
<th>Marker at highest score</th>
<th>Score</th>
<th>Population</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBD7</td>
<td>1p13.1</td>
<td>D15S252</td>
<td>NPL 2.61</td>
<td>Flemish IBD</td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>IBD7</td>
<td>1p33.3-32.3</td>
<td>D15S197</td>
<td>NPL 2.07</td>
<td>Flemish IBD</td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>IBD7</td>
<td>1p36.13-36.11</td>
<td>D18S552-D18S234</td>
<td>LOD 2.64</td>
<td>Non-Jewish American</td>
<td>(Cho et al. 1998)</td>
</tr>
<tr>
<td>1q21.3</td>
<td>D15S305</td>
<td>NPL 2.97</td>
<td>LOD 2.08</td>
<td>North European IBD</td>
<td>(Hampe et al. 1999)</td>
</tr>
<tr>
<td>1q43-q44</td>
<td>D15S2670-D15S2682</td>
<td>LOD 2.08</td>
<td>LOD 1.25</td>
<td>American IBD</td>
<td>(Duerr et al. 2000)</td>
</tr>
<tr>
<td>2q32.3</td>
<td>D2S117</td>
<td>LOD 2.4</td>
<td>LOD 2.4</td>
<td>Canadian IBD</td>
<td>(Rioux et al. 2000)</td>
</tr>
<tr>
<td>IBD9</td>
<td>3p14.2-14.1</td>
<td>D15S1766-D15S285</td>
<td>LOD 2.69</td>
<td>North European IBD</td>
<td>(Satsangi et al. 1996b)</td>
</tr>
<tr>
<td>IBD9</td>
<td>3p21.31</td>
<td>D15S157</td>
<td>LOD 1.31</td>
<td>American IBD</td>
<td>(Duerr et al. 2000)</td>
</tr>
<tr>
<td>3q13.12</td>
<td>D15S3045</td>
<td>LOD 2.3</td>
<td>LOD 2.3</td>
<td>American IBD</td>
<td>(Cho et al. 1998)</td>
</tr>
<tr>
<td>3q25.1-26.31</td>
<td>D18S1279-D18S375</td>
<td>LOD 2.1</td>
<td>LOD 3.23</td>
<td>American IBD</td>
<td>(Hampe et al. 1999)</td>
</tr>
<tr>
<td>3q26.31</td>
<td>D18S3053-D18S2427</td>
<td>LOD 2.07</td>
<td>LOD 2.17</td>
<td>Finnish IBD</td>
<td>(Paavola-Sakki et al. 2003)</td>
</tr>
<tr>
<td>4q25</td>
<td>D4S406</td>
<td>NPL 2.1</td>
<td>LOD 3.0</td>
<td>American IBD</td>
<td>(Cho et al. 1998)</td>
</tr>
<tr>
<td>5q14.1</td>
<td>D5S1501</td>
<td>LOD 1.69</td>
<td>LOD 1.69</td>
<td>Caucasian CD</td>
<td>(Barmada et al. 2004)</td>
</tr>
<tr>
<td>IBD5</td>
<td>5q33.35</td>
<td>-</td>
<td>LOD 2.2</td>
<td>Jewish American CD</td>
<td>(Ma et al. 1999)</td>
</tr>
<tr>
<td>IBD3</td>
<td>6p</td>
<td>D6S197</td>
<td>LOD 3.06</td>
<td>North European CD</td>
<td>(Shaw et al. 2003)</td>
</tr>
<tr>
<td>IBD3</td>
<td>6p22.2</td>
<td>D6S2439</td>
<td>LOD 2.6</td>
<td>Caucasian CD</td>
<td>(Barmada et al. 2004)</td>
</tr>
<tr>
<td>IBD3</td>
<td>6p22.2-21.2</td>
<td>D6S1281-D6S1019</td>
<td>LOD 2.3</td>
<td>Canadian IBD</td>
<td>(Rioux et al. 2000)</td>
</tr>
<tr>
<td>IBD3</td>
<td>6p23.22-23.2-22.2</td>
<td>D6S289-D6S276</td>
<td>LOD 2.07</td>
<td>North European IBD</td>
<td>(Hampe et al. 1999)</td>
</tr>
<tr>
<td>6q24.1</td>
<td>D6S314</td>
<td>NPL 2.44</td>
<td>LOD 2.21</td>
<td>Caucasian IBD</td>
<td>(Barmada et al. 2004)</td>
</tr>
<tr>
<td>6q25.2-26</td>
<td>D6S2436-D6S305</td>
<td>LOD 2.21</td>
<td>LOD 3.08</td>
<td>North European IBD</td>
<td>(Satsangi et al. 1996)</td>
</tr>
<tr>
<td>7q21.11</td>
<td>D7S669</td>
<td>LOD 0.91</td>
<td>LOD 1.57</td>
<td>Caucasian IBD</td>
<td>(Barmada et al. 2004)</td>
</tr>
<tr>
<td>7q21.1-31.33</td>
<td>D7S40-D7S648</td>
<td>LOD 0.91</td>
<td>LOD 1.82</td>
<td>Caucasian IBD</td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>8q12.1-13.1</td>
<td>D8S1113-D8S1136</td>
<td>LOD 1.82</td>
<td>LOD 2.02</td>
<td>Caucasian IBD</td>
<td>(Barmada et al. 2004)</td>
</tr>
<tr>
<td>IBD2</td>
<td>12q21.1-21.2</td>
<td>D12S303-D12S326</td>
<td>LOD 1.82</td>
<td>North European CD</td>
<td>(Hampe et al. 1999)</td>
</tr>
<tr>
<td>IBD2</td>
<td>12q23.3</td>
<td>D12S78</td>
<td>LOD 2.34</td>
<td>Finnish CD</td>
<td>(Paavola-Sakki et al. 2003)</td>
</tr>
<tr>
<td>IBD4</td>
<td>14q11.2</td>
<td>-</td>
<td>LOD 8.2</td>
<td>American CD</td>
<td>(Ma et al. 1999)</td>
</tr>
<tr>
<td>IBD4</td>
<td>14q11.2</td>
<td>D14S261</td>
<td>LOD 3.6</td>
<td>American CD</td>
<td>(Duerr et al. 2000)</td>
</tr>
<tr>
<td>IBD4</td>
<td>14q11.2</td>
<td>D14S50</td>
<td>LOD 2.34</td>
<td>American IBD</td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>IBD4</td>
<td>14q11.2</td>
<td>D14S80</td>
<td>LOD 2.04</td>
<td>American IBD</td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>IBD4</td>
<td>14q13.1</td>
<td>D14S49</td>
<td>LOD 2.44</td>
<td>American IBD</td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>15q26.1</td>
<td>D15S652</td>
<td>LOD 1.82</td>
<td>LOD 2.02</td>
<td>Caucasian IBD</td>
<td>(Barmada et al. 2004)</td>
</tr>
<tr>
<td>15q26.1-26.2</td>
<td>D15S652-D15S3611</td>
<td>LOD 1.82</td>
<td>LOD 2.02</td>
<td>Caucasian IBD</td>
<td>(Barmada et al. 2004)</td>
</tr>
<tr>
<td>IBD8</td>
<td>16p13.13</td>
<td>D16S748</td>
<td>NPL 2.49</td>
<td>American CD</td>
<td>(Cho et al. 1998)</td>
</tr>
<tr>
<td>IBD1</td>
<td>16q12.1</td>
<td>D16S411-D16S419</td>
<td>LOD 5.79</td>
<td>multiscience (IBD)</td>
<td>(Cavanaugh et al. 2001)</td>
</tr>
<tr>
<td>IBD1</td>
<td>16q12.1</td>
<td>D16S409-D16S411</td>
<td>LOD 1.71</td>
<td>European CD</td>
<td>(Hampe et al. 1999)</td>
</tr>
<tr>
<td>IBD1</td>
<td>16q12.1-12.2</td>
<td>D16S409-D16S419</td>
<td>LOD 1.71</td>
<td>European CD</td>
<td>(Hampe et al. 1999)</td>
</tr>
<tr>
<td>18p11.31</td>
<td>D18S62</td>
<td>NPL 2.3</td>
<td>LOD 1.15</td>
<td>American IBD</td>
<td>(Van Heel et al. 2000)</td>
</tr>
<tr>
<td>18q22.2</td>
<td>D18S61</td>
<td>LOD 1.15</td>
<td>LOD 2.0</td>
<td>North European IBD</td>
<td>(Hampe et al. 1999)</td>
</tr>
<tr>
<td>IBD6</td>
<td>19p13.3</td>
<td>D19S591-GATA21G05</td>
<td>LOD 4.6</td>
<td>Canadian CD</td>
<td>(Rioux et al. 2000)</td>
</tr>
<tr>
<td>IBD6</td>
<td>19p13.3-13.2</td>
<td>D19S1034-D19S586</td>
<td>LOD 1.38</td>
<td>American CD</td>
<td>(Cho et al. 1998)</td>
</tr>
<tr>
<td>19q13.31</td>
<td>D19S217</td>
<td>LOD 2.9</td>
<td>LOD 2.9</td>
<td>North European CD</td>
<td>(Van Heel et al. 2003)</td>
</tr>
<tr>
<td>20p12.3</td>
<td>D20S192</td>
<td>NPL 2.6</td>
<td>LOD 1.52</td>
<td>North European IBD</td>
<td>(Hampe et al. 1999)</td>
</tr>
<tr>
<td>22q11.23-12.1</td>
<td>D22S315-D22S421</td>
<td>LOD 1.52</td>
<td>LOD 1.52</td>
<td>North European IBD</td>
<td>(Barmada et al. 2004)</td>
</tr>
<tr>
<td>22q12.1</td>
<td>D22S689</td>
<td>NPL 1.9</td>
<td>NPL 1.9</td>
<td>Flemish IBD</td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>22q13.31</td>
<td>D22S274</td>
<td>NPL 1.7</td>
<td>NPL 1.7</td>
<td>Flemish IBD</td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>Xp21.2-21.3</td>
<td>DXS1202-DXS1214</td>
<td>LOD 2.0</td>
<td>LOD 2.0</td>
<td>North European CD</td>
<td>(Hampe et al. 1999)</td>
</tr>
<tr>
<td>Xp22.11-21.2</td>
<td>DXS1226-DXS1214</td>
<td>LOD 1.59</td>
<td>LOD 1.59</td>
<td>North European CD</td>
<td>(Van Heel et al. 2003)</td>
</tr>
<tr>
<td>Xq28</td>
<td>DXS99</td>
<td>LOD 1.7</td>
<td>LOD 1.7</td>
<td>Flemish IBD</td>
<td>(Vermeire et al. 2004)</td>
</tr>
</tbody>
</table>

LOD: logarithm of the odds; NPL: non-parametric LOD score
is now a crucial step in understanding the contribution and interaction of different gene products in disease onset and pathology. Typically, genome-wide scans identify large loci, spanning several hundreds of genes. Fine mapping of these loci is subsequently used to finally end up with an amenable amount of genes for further study. The latter approach has been successfully employed for the IBD1 locus (16q12), which has resulted in the identification of the first major susceptibility gene for CD, CARD15 (see next paragraph).

The ultimate goal of characterizing new susceptibility genes is a more accurate prediction of disease, which might permit the use of more specific therapy adapted to an individual’s genotype. Furthermore, the identification of these variants offers the possibility to stratify genome-wide linkage studies, increasing the power of identification of new loci (Shaw et al. 2003; van Heel et al. 2003).

**SUSCEPTIBILITY GENES AND DETERMINANTS FOR CD**

The majority of early studies of genetic susceptibility in CD used case-control association studies, in which allele frequencies of candidate genes are compared between patients and healthy controls. However, such studies have been criticized, basically because of the lack of reproducibility: original association studies often indicate a much stronger effect than subsequent ones. The reasons for failing to replicate disease associations have been discussed elsewhere (Colhoun et al. 2003). On the other hand, positional cloning has led to the identification of major genetic variants associated with CD on chromosome 16 (IBD1 locus, CARD15) (Hugot et al. 2001), chromosome 10 (DLG5) (Stoll et al. 2004) and chromosome 5 (IBD5 locus, 5q31 risk haplotype) (Rioux et al. 2001). These genes will be discussed below.

**CARD15**

The IBD1 locus represents the best replicated region, showing linkage to CD and specifically not to UC. It was originally mapped in 1996 (Hugot et al. 1996), and has since been replicated in many studies (Table 1) (Ohmen et al. 1996; Cho et al. 1997; Brant et al. 1998; Cavanaugh et al. 1998; Curran et al. 1998; Annese et al. 1999; Hampe et al. 1999; Akolkar et al. 2001). Furthermore, an international collaborative study on IBD reported a remarkable high linkage score for CD at position D16S411 (Cavanaugh 2001). In 2001, two groups simultaneously identified NOD2 (nucleotide oligomerization domain 2), now officially called CARD15 (caspase activation and recruitment domain 15), as the first susceptibility gene for CD (Hugot et al. 2001; Ogura et al. 2001a). Hugot and colleagues employed the positional cloning strategy while Ogura
Chapter 1

and co-workers identified *CARD15* by the positional candidate gene approach. European and North American patients with CD, including those without a family history, are more likely to have variants in *CARD15* than those without CD. It must, however, be noted that another gene within *IBD1* might additionally be responsible for the linkage with this region. This became even more clear when a genome-wide linkage study was stratified with the *CARD15* variants, and linkage to chromosome 16 was observed in CD patients ~25 centimorgan q telomeric of *CARD15* (van Heel et al. 2003).

A recent study comparing a European and Korean cohort for the common *CARD15* variants illustrated that these variants do not exist in the Korean cohort. Moreover, in this study, no association was found between other SNPs in *CARD15* and CD (Croucher et al. 2003). This implicates that the CD-associated polymorphisms arose after the European/Asian separation. Similarly, no *CARD15* mutations were found in the Chinese (Leong et al. 2003) and Japanese (Inoue et al. 2002) population, and the association of *CARD15* variants is not as strong among the Finns (Helio et al. 2003), Irish (Bairead et al. 2003), New Zealands (Leung et al. 2005) and the Scottish (Arnott et al. 2004). Hence, although ethnically divergent populations may present identical phenotypes, they do not necessarily share the same set of predisposing genes.

**Polymorphisms in *CARD15* associated with CD.** The *CARD15* protein belongs to the NOD1/Apaf-1 family, which comprises cytosolic proteins that are composed of an N-terminal caspase recruitment domain (CARD), a centrally located nucleotide-binding domain (NBD), and a C-terminal leucine-rich regulatory (LRR) domain (Bertin et al. 1999; Inohara et al. 1999; Ogura et al. 2001b). Three common single nucleotide polymorphisms (SNPs) in *CARD15* were independently associated with CD (Figure 2): two missense mutations R702W (c.2104C>T, SNP8) and G908R (c.2722G>C, SNP12), and one frameshift mutation 1007fs (c.3020insC, SNP13), which truncates the protein with 30 amino acids (Hugot et al. 2001). All three variants alter the C-terminal third of the protein, and are within or close to the LRR domain, which is involved in ligand recognition. The heterozygous carrier frequency of these variants within CD ranges from 30 to 50%, while 3 to 15% is homozygous or compound heterozygote, meaning that they carry two variants (Hugot et al. 2001; Ogura et al. 2001a; Abreu et al. 2002; Ahmad et al. 2002; Hampe et al. 2002). By comparison, 8 to 15% of healthy controls are heterozygous and 0 to 1% carries a homozygous variant. The relative risk of developing CD if someone carries one of these variants is increased by a factor of 1.5 to 3 for heterozygous carriers, but the risk increases with a factor of 20 to 40 if someone is homozygous or compound heterozygous. Although this relative risk seems high, it must be stated that the absolute risk for developing CD
Figure 2 The CARD15 protein and mutations that are associated with Crohn's disease (CD) (A) Schematic representation of the functional domains in CARD15 and the three CD-associated mutations SNP8, SNP12 and SNP13. (B) Homology model of the leucine-rich regulatory domain. Two loss-of-function amino acid substitutions are found in this region, and are shown as red sticks. The model was constructed using the Swiss Model web server (Schwede et al. 2003) with the crystal structures of 5 LRR containing proteins (PDB IDs 1DFJ, 2BNH, 2BEX, 1Z7X, and 1A4Y) as template structures. The highly ordered organization of the β-sheets provide the interaction interface of CARD15 with specific bacterial motifs.
is no more than 1 in 25 for homozygous carriers. This reduced penetrance can undoubtedly be explained by the requisite of environmental risk factors and/or additional genetic determinants.

A number of recent reports indicate that common genetic variation in the human genome exists as discrete haplotype blocks, each with a limited diversity. The haplotype structure in a 177-kb region surrounding the \textit{CARD15} gene has been determined (Vermeire \textit{et al.} 2002). A specific haplotype was association with CD, and the three CD-associated SNPs were all unique subvariants of this haplotype.

Different mutations in \textit{CARD15} have been associated with increased susceptibility to psoriatic arthritis (Rahman \textit{et al.} 2003), and to Blau syndrome (Miceli-Richard \textit{et al.} 2001), a rare autosomal dominant disorder characterized by early-onset granulomatous arthritis, uveitis, and skin rash with camptodactyly. Interestingly, the mutations found in Blau syndrome are located in the NBD, suggesting a different molecular pathogenesis between Blau syndrome and CD.

\textit{Functional studies on CARD15 and implications for CD}. The Nod1/Apaf-1 family of proteins displays striking similarity to a class of disease resistance (R) proteins found in plants. Following specific recognition of pathogen products, these R proteins mediate a defence response associated with metabolic alterations and localized cell death at the site of pathogen invasion. The LRR domains of R proteins are highly diverse and appear to be involved in the recognition of a wide array of pathogen components (Dixon \textit{et al.} 2000). Similar to the R proteins, CARD15 appears to play an important role in innate and acquired immunity as a sensor of bacterial components. Specifically, CARD15 participates in the signalling events triggered by host recognition of specific bacterial motifs, and subsequently activates NФxB, the key mediator in the production of pro-inflammatory mediators (Inohara and Nunez 2003). Naturally occurring peptidoglycan (PG) fragments were identified as the microbial motifs sensed by CARD15, more specifically muramyl dipeptide (MDP, GlcNAc-MurNac), found in Gram-negative and Gram-positive bacterial PGs (Girardin \textit{et al.} 2003). However, direct binding of MDP to CARD15 has not yet been demonstrated.

The expression of CARD15 was first thought to be restricted to cells of the myeloid lineage, primarily to monocytes (Ogura \textit{et al.} 2001b; Gutierrez \textit{et al.} 2002). With more accurate techniques, however, expression of CARD15 was shown in epithelial cells (Hisamatsu \textit{et al.} 2003; Rosenstiel \textit{et al.} 2003), keratinocytes (Voss \textit{et al.} 2005), vascular endothelial cells (Oh \textit{et al.} 2005) and paneth cells (Lala \textit{et al.} 2003; Ogura \textit{et al.} 2003). Moreover, its expression is enhanced by pro-inflammatory cytokines and bacterial components via NФxB, a mechanism that may contribute to the amplification of the innate immune response (Gutierrez \textit{et al.} 2002;
Rosenstiel et al. 2003). Consistent with this observation, an elevated expression of CARD15 has been shown in inflamed areas of colonic tissue of CD patients (Berrebi et al. 2003). Although the function of CARD15 in bacterial sensing is widely accepted, its physiological function is less well understood. Consequently, the implications of the CD-associated mutations in disease onset and progression remain unclear. Taken together, five hypotheses have been postulated, and they involve both loss-of-function and gain-of-function of the CARD15 mutations, although the gain-of-function hypotheses have been criticized (Figure 3). The gain-of-function hypotheses evolved from CARD15 knockout and transgenic mice, but are not consistent with what is seen in humans. It is important to note that these hypotheses are not mutually exclusive, and may be physiologically relevant in combination.

In vitro and ex vivo experiments indicated that the three CD-associated polymorphisms actually decreased the activation of NFκB and pro-inflammatory cytokine production, which is inconsistent with the observation that NFκB is up-regulated in patients (Ogura et al. 2001b; Bonen et al. 2003; Hisamatsu et al. 2003). However, this might reflect a lack of triggering a primary innate immune response to bacterial invasion. Since the three CD-associated variants of CARD15 are located in or near the LRR (Figure 2), it was suggested that bacterial sensing is impaired, thus explaining the susceptibility to disease. As a consequence, clearing of bacterial products is inefficient, which might lead to a secondary, compensatory activation of NFκB independent from CARD15 (Figure 3A). Notably, peripheral blood mononuclear cells (PBMC) from individuals homozygous for the major disease-associated SNP13 mutation did not respond to synthetic MDP (Chamaillard et al. 2003; Inohara et al. 2003; Inohara et al. 2005). They also show a defective pro-inflammatory cytokine release after stimulation of mononuclear cells with MDP (Li et al. 2004; Netea et al. 2004; Netea et al. 2005). We have recently shown that PBMC isolated from CD patients carrying CARD15 polymorphisms, produced significantly less IL1β, IL6 and IL10 after stimulation with adherent-invasive E. coli LF82 in a gene-dose effect (Peeters et al. 2006). This was the first study where aberrations were found in heterozygous carriers of SNP8 and SNP12 CARD15 mutations, which is the largest group of patients. Moreover, CARD15 1007fs carriage was also associated in a gene-dose-dependent manner with low mononuclear cell TNF release by stimulation with a combination of interferon gamma (IFNG) and GM-CSF (Halme et al. 2004).

The finding that CARD15 is expressed in paneth cells is particularly of interest, because these cells provide the host defence against microbes in the ileum, while they are not present in the normal colon. Paneth cells secrete a number of antibacterial substances in the lumen of the crypt,
like lysozyme, phospholipase A and α- and β-defensins (Ouellette and Bevins 2001). Lysosyme is an enzyme that breaks down bacterial cell wall components into MDP, which is in turn recognized by CARD15. Moreover, the ileum is an important pathological site in CD, and CARD15 mutations are primarily present in patients with ileal involvement (Cuthbert et al. 2002). Ileal expression of the α-defensins was diminished in active regions in patients with CARD15 mutations, while this was not the case in the diseased colon (Wehkamp et al. 2004). CARD15 is also an inducer of β-defensins, an effect that is lost in CARD15 SNP13 homozygous patients, leading to a defective epithelial defence, proliferation of bacteria and potential loss of epithelial barrier function (Figure 3B).

In human monocyte-derived dendritic cell cultures, CARD15 agonists synergistically induce IL12 production in combination with Toll-like receptor 3 (TLR3), TLR4, and TLR9 agonists to induce T helper type 1 (Th1)-lineage immune responses (Tada et al. 2005). This synergistic effect was lost in patients carrying a mutant CARD15 protein (van Heel et al. 2005b; Kramer et al. 2006). The inflammatory phenotype of CD is difficult to reconcile with a decrease in TNF and IL12 (Sartor 1994). However, it was recently demonstrated that the synergistic induction of IL10 in response to MDP and TLR-stimuli was lost in CARD15 mutant monocyte-derived dendritic cells (Kramer et al. 2006). Interleukin 10 is crucially involved in the down-regulation of the inflammatory process. It was thus postulated that the IL10 mediated immune suppression is impaired and the counter-effect for pro-inflammatory cytokines is lost, thus contributing to chronic ongoing inflammation in CD (Figure 3C).

Incubation of normal murine macrophages with MDP was shown to suppress IL12 secretion induced by stimulation with TLR2 ligands (Watanabe et al. 2004). This suppression did not occur in cells lacking CARD15 or cells expressing a mutant form of CARD15 in transfection experiments. Once secreted, IL12 promotes IFNG and growth and differentiation of Th1 cells (Figure 3D). A major concern however is the reproducibility of these results (Kobayashi et al. 2005). Furthermore, CARD15 and TLR2 stimulation of human PBMC isolated from patients with the 1007fs mutation, led to a loss of synergistic induction of pro-inflammatory cytokines, which is also inconsistent with a TLR2 inhibitory function of CARD15 (van Heel et al. 2005a).

Macrophages of mice carrying a mutant CARD15 equivalent to the 1007fs mutation, were shown to secrete higher levels of the mature form of IL1β upon stimulation with MDP (Maeda et al. 2005). This suggests that the variant CARD15 protein promotes processing of proIL1β to the mature IL1β. Given that IL1β-converting enzyme (ICE) is critical for this processing, it is plausible to assume that mutant CARD15 can activate ICE in response to MDP (Figure 3E).
Figure 3 Loss-of-function hypotheses about the pathogenic mechanisms of CD-associated CARD15 variants (lower panel), compared with normal functions of wild type CARD15 (upper panel). See text p. 17.
Figure 3 Gain-of-function hypotheses about the pathogenic mechanisms of CD-associated CARD15 variants (lower panel), compared with normal functions of wild type CARD15 (upper panel). See text p. 18.
However, macrophages from patients with mutant \textit{CARD15} were shown to have reduced IL1β release upon MDP stimulation (Li \textit{et al.} 2004). The last two observations may indicate important differences in the function of \textit{CARD15} in murine versus human cells.

Recently, the pathway for \textit{CARD15} signalling has been studied. Upon binding of MDP, \textit{CARD15} interacts with RIPK2 (receptor-interacting serine-threonine kinase 2), a kinase that phosphorylates the inhibitor of NFκB kinase, through homophilic CARD-CARD interaction, leading to the nuclear translocation of NFκB (Kobayashi \textit{et al.} 2002). \textit{CARD15} activation also leads to ubiquitylation of the NFκB essential modulator (NEMO) (Abbott \textit{et al.} 2004), and has also been shown to interact with erbb2 interacting protein (ERBIN) (Chen \textit{et al.} 2004) and GRIM-19, a protein with homology to the NADPH dehydrogenase complex (Barnich \textit{et al.} 2005b), both essential for NFκB activation. It was shown that membrane targeting, which is crucial for NFκB activation is impaired in the presence of a SNP13 mutation (Barnich \textit{et al.} 2005a). Two leucine residues and a tryptophan-containing motif in the C-terminal domain of \textit{CARD15} mediate this membrane interaction.

The association of \textit{CARD15} with CD may provide a unifying explanation for several factors influencing the development of this disease. Firstly, \textit{CARD15} responds to bacterial cell wall components, linking intra-luminal bacteria to CD. Secondly, mutations of the LRR domain affect its sensing function leading to aberrant activation of the NFκB pathway, which is abnormally activated in CD (Schreiber \textit{et al.} 1998; Tanabe \textit{et al.} 2004). Thirdly, several studies have indicated that mutations of the \textit{CARD15} gene are associated with ileal involvement (Ahmad \textit{et al.} 2002; Cuthbert \textit{et al.} 2002; Lesage \textit{et al.} 2002), where \textit{CARD15} is highly expressed. Thus, \textit{CARD15} seems to be important for the pathogenesis of CD.

Together with association of CD with \textit{TLR4} and \textit{CD14} (see p.23-25), the involvement of \textit{CARD15} demonstrates the importance of the innate immune response in the pathogenesis of CD, which was discussed in the introduction.

\textbf{DLG5}

IBD-associated variants responsible for the linkage observed at 10q23 (Hampe \textit{et al.} 1999) were attributed to the disk large gene 5 (\textit{DLG5}) by positional cloning (Stoll \textit{et al.} 2004).

Association with the IBD phenotype was strongest, and it was higher in the CD subgroup as compared to the UC subgroup, although this could be ascribed to the smaller UC sample size in this study. This association has since been replicated in three studies (Daly \textit{et al.} 2005; Newman \textit{et al.} 2006; Tenesa \textit{et al.} 2006), while two other studies could not confirm these results (Noble \textit{et al.} 2006).
Two distinct haplotypes were identified with a distortion of transmission in trios. One of the risk-associated \textit{DLG5} haplotypes was distinguished from the common haplotype by a nonsynonymous SNP c.113G>A, resulting in the amino acid substitution R30Q. This SNP was significantly associated with IBD, as was another nonsynonymous SNP c.4136C>A resulting in a P1371Q substitution. Furthermore, a significant difference in association of the 113A variant with CD was observed in affected individuals carrying the risk associated \textit{CARD15} alleles. This suggests a complex pattern of gene-gene interaction between \textit{DLG5} and \textit{CARD15}.

\textit{DLG5} was first cloned in 1998 as a homolog of the drosophila DLG by searching expressed sequence tag databases for related sequences (Nakamura \textit{et al.} 1998). It encodes a scaffolding protein essential for the maintenance of epithelial integrity (Humbert \textit{et al.} 2003). \textit{DLG5} co-localizes to vinexin, involved in cytoskeletal organization and signal transduction, and to β-catenin, a major adherens junction protein (Wakabayashi \textit{et al.} 2003). The expression of \textit{DLG5} has been shown in the intestine and isolated epithelial cells, and thus it seems plausible that mutations in this gene might disrupt epithelial barrier function of the intestine (Stoll \textit{et al.} 2004). \textit{In silico} analysis of the R30Q and P1371Q suggested that both variants probably impair scaffolding functions of \textit{DLG5}.

\textbf{OCTN1 and OCTN2}

A locus of approximately 250 kb at 5q31 (\textit{IBD5}) was previously associated with CD (Ma \textit{et al.} 1999; Rioux \textit{et al.} 2000; Giallourakis \textit{et al.} 2003; Negoro \textit{et al.} 2003). It was an interesting region because it harbours a cytokine cluster (Rioux \textit{et al.} 2001). Peltekova and co-workers re-sequenced five genes included within this region, and reported two variants in the organic cation transporter gene cluster (\textit{OCTN}): a missense substitution in \textit{SLC22A4}, coding for OCTN1 (c.1672C>T, L503F) and a G>C transversion in the promoter of \textit{SLC22A5}, coding for OCTN2 (c.-207G>C) (Peltekova \textit{et al.} 2004). They are both part of a haplotype associated with CD. Moreover, the risk for disease was greater in the presence of both the \textit{IBD5} haplotype and the CD-associated \textit{CARD15} alleles. These observations have since been repeated twice, and genotype-phenotype analysis revealed an association particularly with colonic disease (Torok \textit{et al.} 2005; Waller \textit{et al.} 2005). However, a study from Noble and co-workers could not find an association between the \textit{OCTN1/2} variants and CD in the absence of the \textit{IBD5} haplotype (Noble \textit{et al.} 2005a), thus a causative role for these genes remains plausible but is not yet proven. Further genetic and functional data are required to fully designate these genes as susceptibility genes. In a recent study, an association was found between the promoter polymorphism in
SLC22A5 and psoriatic arthritis (Ho et al. 2005), another chronic inflammatory disease that previously has been associated with CARD15 (Rahman et al. 2003).

OCTN1 and OCTN2 are transmembrane proteins that regulate carnitine and cation transport (Burckhardt and Wolff 2000). They are expressed in cell types that are affected in CD, including epithelial cells, macrophages and T cells (Peltekova et al. 2004). The L503F variant is located within the transport domain, and thus alters its transport function: carnitine uptake in fibroblasts was lower in the L503F variant, while tetraethyl ammonium uptake was increased. The promoter polymorphism in SLC22A5 impaired the binding of the heat shock transcription factor 1. Consequently, the expression of c.-207C OCTN2 was less induced in response to heat shock. These observations could be consistent with an altered uptake of physiological compounds or bacterial metabolites.

Other gene associations for CD

Many genes have been tested for association to CD or IBD in general, because they are positional or functional candidate genes (Table 2). We describe three genes that have been replicated in association studies (TLR4, CD14 and TNF), because of their regulatory role in innate immune responses and functional relevance in IBD pathogenesis.

TLR4. Lipopolysaccharide (LPS), a cell wall component of Gram-negative bacteria, is a major inducer of inflammation, and its signalling is mediated through the cell surface TLR4. During intestinal inflammation, TLR4 is up-regulated on epithelial cells, macrophages and dendritic cells, thus providing a first line defence against enteric Gram-negative bacteria. An association between a polymorphism in the LRR region of TLR4 (c.896A>G, D299G) has been reported within a Dutch CD and UC cohort (Franchimont et al. 2004). Allele frequencies of 10.9% were found in CD patients, versus 5% in healthy controls. The association was replicated twice (Braat et al. 2005; Gazouli et al. 2005), but could not be reproduced in three other studies (Arnott et al. 2004; Torok et al. 2004; Lakatos et al. 2005). This mutation was previously linked to a decreased bronchial responsiveness to LPS (Arbour et al. 2000), and impairs LPS signaling. However, no functional defect, e.g. cytokine release, LPS recognition, has been attributed to heterozygous carriership of this mutant in CD patients (Erridge et al. 2003; von Aulock et al. 2003). A transcriptome analysis of monocyte-derived dendritic cells, isolated from homozygous CARD15 and TLR4 patients, revealed that a large number of genes are differentially regulated in both groups (Braat et al. 2005). This suggests that mutations in different genes can cause similar effects on gene transcription and can thus result in a similar phenotype.
<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosome</th>
<th>Candidate genes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>associated</strong></td>
</tr>
<tr>
<td><strong>Positional</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBD1</td>
<td>16q</td>
<td>CARD15</td>
</tr>
<tr>
<td>IBD2</td>
<td>13q</td>
<td>KRTTHB6</td>
</tr>
<tr>
<td>IBD3</td>
<td>6p</td>
<td>TNF, ILA, HSPA1B, MLN, TAP2</td>
</tr>
<tr>
<td>IBD4</td>
<td>14q</td>
<td>NFIBHA</td>
</tr>
<tr>
<td>IBD5</td>
<td>5q</td>
<td>CD14, OCTN1, OCTN2, IL4</td>
</tr>
<tr>
<td>IBD6</td>
<td>19p</td>
<td>ICAM1, C3</td>
</tr>
<tr>
<td>IBD7</td>
<td>1p</td>
<td>PTPN22</td>
</tr>
<tr>
<td>IBD8</td>
<td>16p</td>
<td>IL4R, ITGAM, IGS6, CD19</td>
</tr>
<tr>
<td>IBD9</td>
<td>3p</td>
<td>TLR9, MLH1</td>
</tr>
<tr>
<td></td>
<td>1q</td>
<td>EPHX1</td>
</tr>
<tr>
<td></td>
<td>4q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7p</td>
<td>CARD4</td>
</tr>
<tr>
<td></td>
<td>7q</td>
<td>ABCB1, SERPINE1</td>
</tr>
<tr>
<td></td>
<td>10p</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10q</td>
<td>DLG5</td>
</tr>
<tr>
<td></td>
<td>17q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>Functional</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1q</td>
<td>IL10</td>
<td></td>
</tr>
<tr>
<td>2q</td>
<td>IL1, IL1R1, CTLA4, SLC11A1</td>
<td></td>
</tr>
<tr>
<td>4q</td>
<td>TLR1, TLR6</td>
<td>UGT1A7</td>
</tr>
<tr>
<td>4p</td>
<td>TLR2</td>
<td></td>
</tr>
<tr>
<td>8p</td>
<td>TLR4</td>
<td></td>
</tr>
<tr>
<td>9q</td>
<td>IL18, MMP3</td>
<td>GSTP1</td>
</tr>
<tr>
<td>11q</td>
<td>TNFRSF1A</td>
<td></td>
</tr>
<tr>
<td>12p</td>
<td>IL6</td>
<td>CYP1A1</td>
</tr>
<tr>
<td>17q</td>
<td>CCL2</td>
<td>AIRE</td>
</tr>
<tr>
<td>20q</td>
<td>BPI</td>
<td></td>
</tr>
<tr>
<td>21q</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22q</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Introduction to the molecular genetics of Crohn’s disease

CD14. The CD14 gene is located within the IBD5 locus, and encodes two protein forms: a glycosylphosphatidylinositol-anchored membrane protein and a monocyte- or liver-derived soluble serum protein that lacks the anchor. Both molecules are critical for LPS dependent signal transduction (Wright et al. 1990). Klein and co-workers described an increased incidence of CD14 c.-159C>T heterozygous and homozygous mutants in CD patients compared to healthy controls (Klein et al. 2002). This association was replicated in a Greek population (Gazouli et al. 2005), but it was not observed in three other studies (Arnott et al. 2004; Torok et al. 2004; Peters et al. 2005). An interaction between CARD15 and CD14 has been observed (Klein et al. 2003).

Although the expression of CD14 on monocytes is higher in CD as compared to healthy controls, its expression did not correlate to the CD14 genotype (Griga et al. 2005). Similarly, CD14 expression is higher in inflamed intestinal mucosa (Rogler et al. 1997), but the association to the CD14 genotype was not found in this study.

TNF. Tumor necrosis factor alpha (TNF) is a pro-inflammatory cytokine that provides a rapid form of host defence against infection but is fatal in excess. It was a strong candidate gene within IBD3, because TNF levels are increased in the serum, mucosa and stool of IBD patients, while anti-TNF therapy is very efficacious in IBD (Targan et al. 1997; Present et al. 1999). TNF production is under a strong genetic influence (Westendorp et al. 1997). Three SNPs in the promoter of TNF (c.-1031C, c.-863A and c.-857T) were associated with susceptibility to CD in a Japanese population (Negoro et al. 1999). In contrast, none of these SNPs could be associated with IBD in two independent Caucasian populations, while the c.-857C allele was more prevalent in IBD and UC (van Heel et al. 2002). Interestingly, this variant was also associated with CD when they left out the common CARD15 allele carriers, meaning that these genes act independently to confer CD susceptibility. Ex vivo LPS challenging of monocytes resulted in a higher production of TNF in c.-857C homozygous individuals. In addition, the transcription factor OCT1, an inhibitor of NFκB transactivation, specifically binds to the c.-857T variant, and not to the c.-857C.

Pediatric onset, colonic disease and familial aggregation of CD was associated with a frequent polymorphism in the binding site for NFκB in the TNF promoter, c.-863C>A (Levine et al. 2005). Electrophoretic mobility shift assay and transfection experiments demonstrated that MDP exposure stimulates TNF gene transcription, as a result of CARD15-induced NFκB activation (Linderson et al. 2005). When the CD-associated CARD15 1007fs variant was analyzed, induction of TNF promoter activity was found to be defective. Different combinations
of *CARD15* and *TNF* promoter polymorphisms gave rise to distinct TNF transcription levels, which means that *CARD15* and *TNF* promoter polymorphisms interact to exert a functional effect on MDP induced TNF production. This gene-gene interaction may contribute to inter-individual variation in susceptibility to CD. The different findings between these studies may reflect genuine population differences, or it could mean that the specifically tested variants are not directly pathogenic.

**CONCLUSION**

Epidemiologic data strongly suggest that genetic susceptibility is a major contributing factor to CD. Viewing the association studies, many allelic variants are involved, and great discrepancies can be found. Usually, this is explained by the widespread heterogeneity of CD, and the fact that single genes contribute little to the complex phenotype. Most associations of candidate genes could not be reproduced, didn’t aid in understanding pathogenesis, or didn’t facilitate diagnosis. Alternatively, genome-wide screening is less biased, and has been proven very useful with the identification of *IBD1* and *CARD15*, which has been replicated many times. Several aspects of future linkage and association analysis should be taken into account: 1) clinical classifications of individuals should be more conform; 2) sample size is a critical determinant for false-negative errors in such studies, especially when stratifying to phenotype or genotype; 3) association of genes is usually performed with selected variants of the gene, instead of sequencing the whole gene. In addition, knowledge of the haplotype structure surrounding genes of interest, allows for common variation in a gene to be tested with more statistical power, even if the causal variant within the shared haplotype block has not been identified (Daly *et al.* 2001).

In 1997, a group of scientists and clinicians founded the IBD International Genetic Consortium (IBDIGC), now a group of twelve research groups involved in the study of genes that are implicated in IBD (Cavanaugh 2003). They collaboratively study large numbers of well-documented families for linkage. Similarly, small-scale pooling of populations may significantly increase the power of linkage and association studies. Finally, the correct choice of candidate genes is another major issue. The IBD loci that have been repeatedly verified are to be selected preferentially, especially if functional data regarding the gene is available through other analysis such as expression data.

**REFERENCES**


Introduction to the molecular genetics of Crohn’s disease


Chapter 1

the Greek population." World J Gastroenterol. 11(5): 681-5.


Introduction to the molecular genetics of Crohn's disease


Introduction to the molecular genetics of Crohn’s disease


Introduction to the molecular genetics of Crohn’s disease


Lessons from epidemiology." Dig Dis 21(2): 91-104.


Chapter 2

A FUNCTIONAL PROMOTER POLYMORPHISM IN CARD15

Debby Laukens¹, Michaela Loos¹, Jurgen Del-Favero², Sara Bogaert², Harald Peeters³, Martine De Vos³ and Pieter Rottiers¹

¹Department of Molecular Biomedical Research, Ghent University and Flanders Interuniversity Institute for Biotechnology (VIB), Gent, Belgium
²Department of Molecular Genetics, University of Antwerp, VIB, Antwerpen, Belgium
³Department of Gastroenterology, Ghent University Hospital, Gent, Belgium

ABSTRACT

Background & aims: Three polymorphisms within the CARD15 gene have been independently associated with Crohn’s disease (CD). Although this association is strong, an average of 60 to 70% of CD patients do not have any of these three mutations. Furthermore, they do not account for all of the linkage observed between CD and the IBD1 locus. Since CARD15 is highly up-regulated during inflammation, we looked for polymorphisms in the CARD15 promoter region.

Methods: The promoter region of CARD15 was sequenced in 28 Flemish CD patients, and a polymorphism was genotyped in 104 CD patients and 60 controls. An in vitro gene reporter assay was used to assess the functionality of the promoter polymorphism in response to tumor necrosis factor alpha (TNF). Public HapMap data from CEPH trios were used to determine the haplotype structure surrounding CARD15.

Results: Sequencing of the promoter region revealed a c.-59G>A polymorphism in the 5’ untranslated region (5’UTR), contained within a conserved potential binding site for the E2F transcription factor. This polymorphism was significantly associated with CD. However, in 99% of the cases, the CD-associated CARD15 mutations occur in the c.-59G>A background, suggesting that they are linked. In addition, the infrequent c.-59A allele was less responsive to TNF stimulation. The complete CARD15 gene is located within a single haplotype block.

Conclusions: A functional promoter polymorphism, linked to the common CD-associated mutations, was identified in the CARD15 5’UTR. The complete CARD15 gene is located in a
single haplotype block, and we are currently determining the haplotype structure in our patient population to evaluate the transmission of the CARD15 promoter polymorphism.

INTRODUCTION

Crohn’s disease (CD, MIM 266600) is an inflammatory condition of the gastrointestinal tract. The disease is believed to result from an interaction between environmental factors and genetic predisposition. Since 1996, a locus at 16q12 has been repeatedly associated with increased susceptibility to CD (Hugot et al. 1996). In 2001, two groups independently identified the first strongly CD-associated gene called NOD2, later referred to as CARD15 (Hugot et al. 2001; Ogura et al. 2001a). CARD15 is a cytosolic receptor for bacterial components and initiates an inflammatory response following bacterial challenge. Both groups found the same gene using a different approach. Hugot and co-workers systematically refined the susceptibility locus of chromosome 16. They characterized CARD15 within this region, and sequenced the 11 exons of this gene in 50 unrelated CD patients. Three haplotypes with preferential transmission to affected individuals were identified, each containing one rare allele of single nucleotide polymorphism (SNP) 8 (c.2104C>T), SNP12 (c.2722G>C) or SNP13 (c.3020insC) in the context of a common background. SNP8 and SNP12 result in non-conservative amino acid substitutions R702W and G908R respectively. SNP13 is a frameshift mutation (1007fs), which truncates the tenth leucine-rich repeat of CARD15. On the other hand, Ogura and colleagues used the candidate gene approach, which is based on an intelligent search for genes within regions of linkage. They identified the SNP13 mutation, which was preferentially transmitted, and assessed this SNP in a large case-control study. The association of the three mutations have since been confirmed in many studies. In Europe, they are found in approximately one third of CD patients, especially in those with ileal disease (Cuthbert et al. 2002).

The observed linkage of CD to chromosome 16 can not be entirely explained by the present associations (Hugot et al. 2001). Furthermore, linkage to chromosome 16 was still observed in CD patients not possessing one of the three CARD15 mutations (Hampe et al. 2002; Shaw et al. 2003; van Heel et al. 2003). Thus, other variants of this gene or variations in neighbouring genes on chromosome 16 may be involved in CD susceptibility.

The expression of the CARD15 protein in normal colon is restricted to scattered mononuclear cells in the lamina propria (Berrebi et al. 2003), and in paneth cells in the ileum (Ogura et al. 2003). In the CD colon, however, the number of positive cells is correlated with the inflammatory infiltrate. In severely inflamed CD colon samples, intestinal epithelial cells were
A promoter polymorphism in CARD15

also CARD15 positive. Together with the observation that CARD15 is up-regulated in acute appendicitis, this favours the hypothesis that CARD15 is induced upon inflammatory stimuli. Furthermore, it was shown that tumor necrosis factor alpha (TNF), a potent pro-inflammatory cytokine, up-regulates CARD15 via NFκB in epithelial cell lines (Rosenstiel et al. 2003). In accordance, two functional NFκB binding sites were found within the CARD15 promoter. Therefore, CARD15 expression regulation could be an important mechanism in innate immune response in intestinal epithelial cells. We wondered whether polymorphisms are present within the promoter of CARD15, which might influence their expression. We sequenced a part of the promoter region in CARD15, and found a promoter polymorphism in the 5' untranslated region (5'UTR), c.-59G>A, located within a DNA stretch that potentially binds the E2F transcription factor. The functionality of both alleles was tested in an in vitro gene reporter system.

MATERIALS AND METHODS

Sequencing and genotyping of c.-59G>A polymorphism
Genomic DNA was extracted from whole blood using the Qiagen blood and cell culture DNA kit. The promoter region of CARD15 was PCR-amplified in 28 Flemish CD patients using forward primer: 5'-GGC CTG TCC CCT CGT GAA TG-3' and reverse primer: 5'-GTC GCG GCC AAG GAT GAA AG-3'. This product was sequenced with the BigDye Terminator v3.1 Cycle Sequencing kit and analyzed on an ABI3700 analyzer (Applied Biosystems).

The c.-59G>A polymorphism was genotyped by RFLP-PCR in 104 unrelated CD patients and 60 healthy controls. A fragment containing the polymorphism was amplified with the following primers: forward: 5'-GGC GGA GGT TGG AGT TGA AAA TAA-3'; reverse: 5'-GGA AGC CAG GAT CTA AGG TA-3'. The PCR product was precipitated with SeeDNA (Amersham Biosciences), digested with BsmI, and restriction fragments were analyzed on a 2% agarose gel. The presence of the c.-59A allele creates a BsmI site, resulting in two bands of 1108 and 510 bp, instead of a single 1618 bp band in the presence of the c.-59G allele.

Genotyping for the three CD-associated CARD15 polymorphisms c.2104C>T, c.2722G>C and c.3020insC was performed using RFLP-PCR as previously described (Laukens et al. 2005).

Constructs
Genomic PCR fragments of 155 bp of the promoter region of CARD15, starting at base -1 to -155 upstream from the initiation codon, were cloned between the SacI and KpnI sites of the PGL3basic luciferase reporter vector (Promega Corporation Benelux). This generated PGL3-
59A and PGL3-59G constructs. Both constructs were sequence verified. The β-galactosidase expressing pUT651 vector (Eurogentec, Belgium) was used for normalizing transfection efficiencies.

**Transfection and gene reporter assay**

HEK293T cells (ATCC CRL1573) were seeded in 24-well plates at 5x10^5 cells per well the day before transfection. Cells were cotransfected with 100 ng pUT651 and 100 ng of PGL3-59G or PGL3-59A in triplicate using the calcium phosphate method. Twenty-four hours posttransfection, cells were incubated with 1000 U/ml recombinant human TNF purified from *E. coli* (produced in-house). After 24 hours, cells were lysed in 150 µl lysis buffer (25 mM Tris pH 7.8, 2 mM DTT, 2mM CDTA, 10% glycerol and 1% Triton X-100) for 15 minutes. Lysates were analyzed for luciferase activity by mixing 50 µl of lysate with 50 µl of luciferase assay buffer (20 mM tricine, 1.07 mM (MgCO_3)_2M(OH)_2, 2.67 mM MgSO_4, 33.3 mM DTT, 0.1 mM EDTA, 270 µM CoA, 530 µM ATP and 470 µM luciferin). Luciferase activity was measured in a TopCount luminescence counter (Packard). β-galactosidase activity was analyzed by mixing 20 µl of lysate with 160 µl of substrate buffer (60 mM Na_2HPO_4 pH 7.0, 10mM KCl, 1mM β-mercaptoethanol) and 20 µl of chlorophenolred-β-D-galactopyranoside (CPRG) substrate (Roche Diagnostics NV, Belgium). After 30 minutes, β-galactosidase activity was measured at 540 nm in a microplate reader (BioRad Laboratories, Belgium).

**Haplotype structure**

Public data of the International HapMap Project (http://www.hapmap.org) were used to identify the haplotype structure surrounding *CARD15*, using the CEPH (Centre d'Etude du Polymorphisme Humain) population data, who are all Utah (USA) residents with ancestry from northern and western Europe. The criteria used to assign membership in the CEPH population have not been specified, except that all donors were residents of Utah.

**Statistics**

Luciferase induction was represented as mean of triplicates + SD. The Student’s t test was used to compare groups. Allele frequencies between controls and CD patients were compared using Pearson’s χ^2 test. Two-tailed probabilities were calculated, and *P*-values of less than 0.05 were considered statistically significant.
RESULTS

A polymorphism in the promoter region of CARD15 is linked to CD-associated mutations

To search for polymorphisms in the promoter of CARD15, a 425 bp region surrounding the most functional NFκB binding site (Rosenstiel et al. 2003) was sequenced in 28 unrelated CD patients. One polymorphism was found, c.-59G>A, and subsequently typed in 104 CD patients and 60 controls (Table 1). An association with CD patients was found for SNP8, SNP12, SNP13 and c.-59G>A, but the latter significance is because it is in linkage disequilibrium (LD) with SNP8 (R²=0.224, LOD 11.76). If we omit the patients carrying a CARD15 mutation, the association of c.-59A with CD is lost. Furthermore, no correlation was found with a specific phenotype of CD, e.g. disease location, age of onset or structuring disease (data not shown).

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Frequencies for SNP8, SNP12, SNP13 and c.-59G&gt;A in our study population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% rare allele frequency in CD patients (N=208)</td>
</tr>
<tr>
<td>SNP8</td>
<td>17.7</td>
</tr>
<tr>
<td>SNP12</td>
<td>3.8</td>
</tr>
<tr>
<td>SNP13</td>
<td>10.1</td>
</tr>
<tr>
<td>c.-59G&gt;A</td>
<td>47.1</td>
</tr>
<tr>
<td>c.-59G&gt;A no CARD15</td>
<td>53.2</td>
</tr>
</tbody>
</table>

The c.-59G>A polymorphism is located within a conserved binding site for E2F

A search for transcription factor binding sites using the TRANSFAC database revealed that the c.-59G>A is located within a potential binding site for the E2F transcription factor (Figure 1). To determine whether this binding site is conserved, a search of E2F binding sites in the murine CARD15 gene and in other members of the Nod1/Apaf1 family (Ogura et al. 2001b) was performed. The human CARD15, NOD1 and the murine CARD15 contain 1 potential binding site for E2F within a 800 bp promoter region upstream of the coding sequence. The Apaf1 promoter contains 2 of these sites. The conservation of the E2F binding region suggests that regulation of CARD15 by E2F might be an essential mechanism in the function of CARD15.
Figure 1 Part of the promoter region of CARD15, containing a functional NFκB site and a potential binding site for E2F. The c.-59G>A polymorphism in the 5’UTR (grey box) is located within the potential binding site for E2F. The E2F consensus sequence is T/C T T C/G G/C G. The coding sequence of exon 1 is boxed.

The c.-59A allele impairs CARD15 induction by tumor necrosis factor alpha
Because the c.-59G>A polymorphism is located within a potential binding site for the E2F transcription factor, we assessed whether the distinct alleles influence the promoter induction after stimulation with tumor necrosis factor alpha (TNF). Therefore, we generated two plasmids containing the c.-59G (PGL3-59G) or c.-59A allele (PGL3-59A), together with the functional NFκB site in a gene reporter system. After 24 hours of TNF stimulation, luciferase activity induction in HEK293T cells transfected with the PGL3-59A construct was significantly less as compared with PGL3-59G construct (Figure 2). A similar result was obtained when HT29 colonic epithelial cells were used as an acceptor cell (data not shown).

CARD15 is contained within a single haplotype block
The haplotype structure surrounding CARD15 was determined using public CEPH data from the HapMap project. The complete CARD15 gene is located within a single haplotype block (Figure 3).

CONCLUSION
We identified a polymorphism within the CARD15 5’UTR that differs in responsiveness to TNF induction. This polymorphism, currently present in the SNP database as rs.5743266, is in LD with SNP8, a common CD-associated polymorphism.
A promoter polymorphism in CARD15

Because the function of the three polymorphisms in CARD15 is not completely understood, and because other genetic variants in or close to CARD15 are supposed to be present, we focused on a promoter polymorphism, which influences the expression of CARD15. Several reports have shown that CARD15 expression is induced by TNF and under inflammatory conditions (Berrebi et al. 2003; Rosenstiel et al. 2003). It is however not known whether the three CD-associated mutations influence CARD15 expression. In a small cohort, there appeared to be no correlation between CARD15 mutations and CARD15 expression in the colon. However, only 4 out of 8 CD patients carried one of the CARD15 mutations (Berrebi et al. 2003). Therefore, a detailed expression study of colonic and ileal CARD15 in relation to the common mutations needs to be performed. Moreover, the importance of CARD15 expression has not been extensively studied. It was shown that overexpression of CARD15 in intestinal epithelial cells leads to a higher IL8 secretion in response to lipopolysaccharide (Rosenstiel et al. 2003). The IL8 cytokine is a chemotactic protein, which recruits neutrophils, and is a major mediator of the inflammatory response. Up-regulation of CARD15 in intestinal inflammation may therefore represent a generic response to bacterial invasion.

The induction of CARD15 transcription in response to inflammation is probably mediated through two functional NFκB sites in the promoter (Rosenstiel et al. 2003). The c.-59G>A polymorphism is located within a potential binding site for the E2F transcription factor. E2F is involved in cell cycle progression (Fang and Han 2006). To establish a true functional consequence of the polymorphism, DNA binding of E2F to both alleles needs to be determined. Because the c.-59A allele is the common background of SNP8, SNP12 and SNP13, the question arises as to the importance of the common mutations in CD. Interestingly, c.-59G>A occurs within the same haplotype block as the CD-associated CARD15 polymorphisms. The precise haplotypes have to be defined. Therefore, we are currently identifying the haplotype structure in
relation to the promoter polymorphism. It was shown that CARD15 with the SNP13 mutation is functionally unable to detect bacterial peptides, resulting in less NFκB activation (Chamaillard et al. 2003; Inohara et al. 2003; Inohara et al. 2005). We here showed that a promoter polymorphism c.-59G>A influences the expression of CARD15 in response to TNF. Perhaps it is the combination of SNP8, SNP12 or SNP13 with the promoter polymorphism that causes susceptibility to CD.

Figure 3 LD across CARD15 and flanking genomic region, based on public CEPH data. $R^2$ values for pairwise LD between each marker are represented (white: $r^2$=0, shades of grey: $0>r^2>1$, black: $r^2$=1). The top of the figure shows distances between markers. Most markers within the CARD15 gene are in strong LD.

In summary, we found a new functional promoter polymorphism in CARD15, but to fully understand its association to CD, detailed haplotype analysis in our patient population is currently ongoing.

ACKNOWLEDGEMENTS

We acknowledge the International HapMap Consortium (HapMap 2003) for the CEPH data, and the Protein expression and Purification core facility of VIB for providing recombinant TNF.
A promoter polymorphism in CARD15

REFERENCES


Chapter 2


DEBRED METALLOTHIONEIN EXPRESSION IN COLONIC CROHN’S DISEASE: EVIDENCE FOR MTF1 AS A NEW DISEASE-MODIFYING GENE

Debby Laukens¹, Harald Peeters², Sara Bogaert², Anouk Waeytens⁴, Bert Vander Cruyssen³, Dirk Elewaut³, Filip De Keyser³, Herman Mielants³, Claude Cuvelier⁴, Eric M. Veys³, Kenny Knecht⁵, Paul Van Hummelen⁶, Jurgen Del-Favero⁷, Erik Remaut¹, Martine De Vos², Pieter Rottiers¹.

¹Department of Molecular Biomedical Research, Ghent University and Flanders Interuniversity Institute for Biotechnology (VIB), Gent, Belgium
²Department of Gastroenterology, Ghent University Hospital, Gent, Belgium
³Department of Rheumatology, Ghent University Hospital, Gent, Belgium
⁴Department of Pathology, Ghent University Hospital, Gent, Belgium
⁵Applied Maths BVBA, Sint-Martens-Latem, Belgium
⁶MicroArray Facility, VIB, Leuven, Belgium
⁷Genetic Service Facility, VIB, Antwerpen, Belgium

ABSTRACT

Background & aims: Crohn’s disease (CD) is a multifactorial disorder involving both genetic and environmental factors. This study aimed at identifying and characterizing new candidate susceptibility genes for CD by integrating known disease loci with gene expression variation in non-inflamed colon biopsies of CD patients.

Methods: Sixteen CD patients and 11 controls were subjected to microarray analysis. Expression of metallothionein (MT) was analyzed in intestine and blood by quantitative PCR and immunohistochemistry. MT-knockdown HT29 cells were generated by small interfering RNA. MT1M, and its transcriptional regulator MRE-binding transcription factor 1 (MTF1), were screened for mutations by sequencing.

Results: Eighteen differentially expressed baseline genes were identified. We focused on the
reduced expression of MT in intestine and blood of CD patients with colonic involvement. We showed that MT induction was not impaired in these patients. To model lowered MT expression in epithelial cells, we created MT-knockdown HT29 cells. These showed a reduced IL8 secretion in response to bacterial challenge, suggesting deficient inflammatory responses in the CD colon. No mutations were found in *MT1M* exons and its promoter region. A polymorphism in *MTF1* (IVS1-128A>T) was associated with disease location. Gene-gene interaction of *MTF1* and *CARD15* predisposes individuals to a high risk for ileal disease.

**Conclusions:** Combining microarray screening with genetic linkage data is an effective tool for identifying novel candidate susceptibility genes. We showed that deficient basal MT expression in CD patients with colonic involvement is genetically determined. The IVS1-128A>T polymorphism in *MTF1* is linked to disease location and serves as a new disease-modifying gene.

**INTRODUCTION**

The strong influence of genetic determinants in Crohn’s disease (CD, MIM 266600) has been shown by familial clustering, and by the high concordance rate in monogenic twins. First-degree relatives of affected individuals show a 20 to 50-fold higher risk for developing CD. Moreover, affected siblings frequently develop the disease at similar ages, and concordance rates reach 80% for disease site, behaviour and presence of extra-intestinal manifestations (Peeters *et al.* 1996; Halfvarson *et al.* 2003). The mode of inheritance of CD is complex, and the number of genes predisposing to CD or modifying its course is currently unknown. However, it is not expected that a single risk gene is sufficient for disease development.

Two complementary approaches are used to identify determinants underlying genetically complex traits such as CD: candidate gene association studies and genetic linkage. In the candidate gene approach, genes that might be involved are screened for polymorphisms in a limited number of patients, and then frequencies in large populations of patients and controls are compared. In genetic linkage, candidate regions are identified by scanning the entire chromosome with microsatellite markers. These so-called disease loci can rarely be narrowed down to less than a few megabases. Nevertheless, if the candidate region contains a manageable number of genes, a candidate gene is chosen on the basis of a known property, e.g. function or tissue-specific expression. Unfortunately, mapping CD loci is hampered by phenotypic variability, genetic heterogeneity across populations, uncontrolled environmental influences, reduced penetrance, and limited statistical power in such studies. Nevertheless, several CD loci
Reduced metallothionein expression in Crohn’s disease

have been corroborated by more than one independent study. The most frequently identified locus for CD is 16q12 (IBD1), which led to the mapping of the first gene to be firmly associated with CD, called CARD15 (Hugot et al. 2001; Ogura et al. 2001). Today, there is evidence for more than 20 loci, but the identity of the causative genes remains largely unknown (Brant and Shugart 2004a). In addition, success in finding susceptibility genes has been limited by the modest effect of individual genes on the complex phenotype.

We applied a complementary approach to identify new potential susceptibility genes. Because much of the variation in gene expression is genetically transmitted (Cheung and Spielman 2002), we evaluated variation of gene expression in unaffected biopsies from CD patients, and combined these results with known linkage data. For this purpose, we looked at gene expression in biopsies taken from non-inflamed regions of the colon from CD patients and healthy controls. We selected genes that show variation in baseline expression and that are also located near a locus for CD. We focused on characterizing one of the candidate genes, metallothionein (MT), and the consequences of its altered expression. This led to the identification of a new disease-modifying gene, MTF1, located at 1p33 (IBD7).

MATERIALS AND METHODS

Patients and biopsies. CD was diagnosed according to clinical, endoscopic and histological criteria. Patients were classified according to the Vienna classification (Gasche et al. 2000). Disease location, however, was defined as the maximal spread of inflammation during the entire follow-up. We thus defined three subgroups: pure colonic involvement (C), both ileal and colonic involvement (IC), and ileal involvement only (I). For RNA extraction, 3 colon biopsies from each of 54 CD patients and 30 controls were collected during colonoscopy. All biopsies were taken from endoscopically normal regions of the sigmoid, immediately placed in RNAlater (Ambion, Cambridgeshire, UK) and stored at -80°C. For immunohistochemical analysis, colon biopsies were collected from 22 CD patients and 8 controls. Additional ileal biopsies were collected from 12 CD patients and 8 controls. Biopsy specimens were immersed in 4% formalin (Labonord, France).

RNA extraction. Total RNA was extracted from biopsies using the RNeasy Mini Kit (Qiagen, Westburg BV, The Netherlands) with on-column DNAsae treatment (Qiagen). Needle homogenization was performed. The quality of RNA used for microarray analysis was checked on the Agilent 2100 Bioanalyzer (Agilent Technologies, Germany). Total RNA from 3 ml of
whole blood was extracted using the RNeasy Midi Kit (Qiagen) following lysis of red blood cells with erythrocyte lysis buffer (Qiagen). Concentration and purity (ratio \( \text{OD}_{260} / \text{OD}_{280} \) between 1.8 and 2.2) of the RNA used for quantitative real-time PCR was checked on a spectrophotometer (UV-1601, Shimadzu Benelux, Belgium).

**Microarray hybridization, scanning and analysis.** The construction of the focus microarray chip, containing 6,779 expressed sequence tags and specifically designed for the study of colonic gene expression, has been described (Laukens *et al.* 2006). Total RNA (5 \( \mu \)g) was amplified using a modified protocol for *in vitro* transcription (Puskas *et al.* 2002). Probe labelling, hybridization, washing and scanning were carried out at the MicroArray Facility (MAF, Leuven, Belgium) of the Flanders Interuniversity Institute for Biotechnology (VIB) as described on http://www.microarrays.be/service.htm. Images were analyzed with ArrayVision (Imaging Research Inc., Canada). Each hybridization was repeated in a dye swap experiment. Spot intensities were measured, corrected for local background, and those that exceeded the background by more than two standard deviations (SD) were included in the analysis. For each gene, ratios of red (Cy-5) to green (Cy-3) intensities (I) were calculated and normalized via a Lowess Fit of the \( \log_2 \) ratios \( \left[ \log_2(\text{ICy-5} / \text{ICy-3}) \right] \) over the \( \log_2 \) total intensity \( \left[ \log_2(\text{ICy-5} \times \text{ICy-3}) \right] \).

A mixture of RNA from 5 CD patients, 5 non-CD inflammatory controls and 5 healthy controls served as reference RNA for comparison of the microarray datasets. This reference sample provides a positive hybridization signal at each probe element on the microarray, which is essential when calculating and comparing fluorescence ratios. The data were imported into GeneMaths\textsuperscript{®} XT (Applied Maths, Belgium). Weighted mean ratios and their corresponding error (pixel SD) were calculated from the dye swap. Data were normalized over all arrays, and missing values were imputed using k-nearest neighbour algorithm (20 neighbours). GeneMaths\textsuperscript{®} XT was used to perform all subsequent analyses.

**Quantitative real-time PCR (qPCR).** One \( \mu \)g of total RNA was converted to single stranded complementary DNA (cDNA) by reverse transcription (Superscript, Gibco, Invitrogen, Belgium) with oligo dT priming. One tenth of the cDNA was used in real-time quantification using the SYBR green kit (Eurogentec, Belgium) and 300 nM of each primer. A two-step program was run on the iCycler (BioRad Laboratories, Belgium). Cycling conditions were 95°C for 10 minutes and 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. Melting curve analysis and direct
sequencing of amplicons on the ABI3700 analyzer (Perkin Elmer, Applied Biosystems, Belgium) confirmed primer specificities. All reactions were run in duplicate and normalized to glyceraldehyde phosphate dehydrogenase (GAPDH) levels. GAPDH was chosen after checking the expression stability of a set of housekeeping genes in biopsies of CD patients and controls using the Genorm software (Vandesompele et al. 2002). Primers were designed using the Beacon Designer software (PREMIER Biosoft International, USA). Sequences of all primer sets are listed in Table 1.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Reference sequence</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>NM_002046</td>
<td>TGC ACC ACC AAC TAC TTA GC</td>
<td>GGC ATG GAC TGT GGT CAT GAG</td>
</tr>
<tr>
<td>MT1A</td>
<td>NM_005946</td>
<td>GCA AAG GGG CAT CAG AAG AGT G</td>
<td>AAA TAC AGT AAA TGG GTC AGG GGT G</td>
</tr>
<tr>
<td>MT1B</td>
<td>NM_005947</td>
<td>AAG TGC TGC TGC TGC TGC TG</td>
<td>TGG TGT CTC TAT TTA TGT CTG CGA G</td>
</tr>
<tr>
<td>MT1E</td>
<td>NM_175617</td>
<td>GCA TCC TCT GGG TCT GGG TTC</td>
<td>AAC AGC AGC CTG GGG AGG AAG</td>
</tr>
<tr>
<td>MT1F</td>
<td>NM_005949</td>
<td>GCG ACT GAT GCC AGG ACA AC</td>
<td>CAC AGG AAA AGG AAT GTA GCA AAT G</td>
</tr>
<tr>
<td>MT1G</td>
<td>NM_005950</td>
<td>GCA AAG GGG CAT CGG AGA AGT G</td>
<td>AAG GGA ATG TAGCAA AGG GGT CAA G</td>
</tr>
<tr>
<td>MT1J</td>
<td>AF348994</td>
<td>GCT GTG CCT GAT GTG GGA AC</td>
<td>AAA TGC AGC AAA TGG CTC AGT ATT G</td>
</tr>
<tr>
<td>MT1M</td>
<td>NM_176870</td>
<td>CTG CAA AGG GAC GTT GGA GAA C</td>
<td>CAG CAA ATG GCT TAT CTA ATT</td>
</tr>
<tr>
<td>MT2A</td>
<td>NM_005953</td>
<td>AAA GGG CGG TCG GAC AAG TG</td>
<td>GAA TAT AGC AAA CGG TCA CGG TCA G</td>
</tr>
<tr>
<td>MTF1</td>
<td>NM_005955</td>
<td>TAA GAC TCA ATT GAT TCA GGG ACG AGA</td>
<td>CAG TTG TGA GAA ATG AAA ACG TAA TGA</td>
</tr>
</tbody>
</table>

**Immunohistochemistry.** Colon and ileum biopsies fixed in formalin were embedded in paraffin according to standard procedures. Sections were rehydrated by serial immersion in xylene and ethanol. Immunostaining was performed on a NexES IHC automated staining system (Ventana Medical Systems, France) using a 1:10 dilution of a mouse monoclonal anti-metallothionein antibody (clone E9, Zymed Laboratories, Sanbio, The Netherlands). An isotype-specific irrelevant antibody (X0931, DakoCytomation, Belgium) was used to control for non-specific binding of the primary antibody. Slides were blindly scored by two pathologists. Semi-quantitative scoring was based on staining intensity and the number of positive cells.

**Induction assays.** The mononuclear cell fraction (PBMC) was isolated from 30 ml of venous blood by density gradient centrifugation in Ficoll-Paque (Amersham Biosciences, The Netherlands). PBMCs were seeded in 6-well plates at 10^6 cells per well in RPMI medium (Gibco, Invitrogen) supplemented with 10% FCS. The next day, they were stimulated with 200 µM ZnSO₄ (Sigma, Belgium), 2 µM dexamethasone (Sigma), 10 ng/ml recombinant human IL6 (Peprotech, Campro Scientific, The Netherlands), 1000 U/ml human recombinant
TNF (produced in-house) or 200 U/ml human recombinant IFNγ (Biosource Europe, Belgium). The cells were lysed in 350 µl RLT buffer (Qiagen) after 3, 6 or 24 hours. To induce oxidative stress, PBMCs were incubated with 100 µM H2O2 (Sigma), which was washed off with PBS after 10 minutes. Cells were lysed after 2, 4 or 6 hours. Total RNA was isolated using the Rneasy Mini Kit (Qiagen), and converted to cDNA for subsequent qPCR. Normalized Ct values were corrected for background at each time point (e.g. dCt\text{induced},t1 - dCt\text{un-induced},t1), and plotted against time (0, 3, 6 and 24 hours, or 0, 2, 4 and 6 hours for H2O2). Induction was calculated as the area under the curve (AUC) for the three time points (AUCs are shaped as triangles and rectangles, and their summation or integration gives the same result):

\[
AUC = \int_0^3 (m_1 \cdot x) dx + \int_3^6 (m_2 \cdot x + b) dx + \int_6^{24} (m_3 \cdot x + c) dx,
\]

where \(m_1\) = the slope of the 0-3 line, \(m_2\) = the slope of the 3-6 line, \(m_3\) = the slope of the 6-24 line, \(b\) = the y-intercept of the 3-6 line, and \(c\) = the y-intercept of the 6-24 line.

HT29 human colon epithelial cells (American Type Culture Collection (ATTC) HTB38) and MT-knockdown cells (HT29MTkd) were seeded in 24-well plates at 10^6 cells per well in RPMI supplemented with 10% FCS. The following day, cells were infected in triplicate at multiplicities of infection (MOI) of 100 with adherent-invasive Escherichia coli strain LF82, isolated from a patient with CD, (a gift from A. Darfeuille-Michaud) (Darfeuille-Michaud et al. 2004) for 1 hour, followed by gentamycin treatment (100 µg/ml) for 1 hour. Supernatants were collected after 8, 24 and 48 hours of infection/induction.

**Small interfering RNA (siRNA).** The short hairpin RNA (shRNA) fused to an H1 promoter was synthesized by PCR on the pSUPER vector (Tronolab, Switzerland), using primers: 5’-CAA TCT CTT GAA TTG CAC TTG CAG GAG CCG GGG GGA TCT GTG GTC TCA TAC AGA ACT TAT AA-3’ and 5’-CCA TCG ATT TCC AAA AAC CGG CTC CTGCAA G TG CAA TCT CTT GAA TTG C-3’. The sequence in bold is a 19-mer that specifically targets MT, and is present in MT1B, MT1E, MT1H, MT1J and MT1M. The PCR fragment was cloned in pLVTH-siGFP (Tronolab, Switzerland). To produce virus for delivery of the shRNA, HEK293T cells (ATCC CRL1573) were transfected by the calcium phosphate method with 3 µg pCMV-d8.91 (Tronolab, Switzerland), 1.5 µg PMDG2 (Tronolab, Switzerland) and 1.5 µg of the shRNA construct. After 48 hours, supernatant containing viral particles was harvested, and put through a 0.45 µm filter. One day before viral transduction, HT29 cells were seeded at 5x10^4 cells per well in a 24-well plate. They were overlaid with viral supernatant, and centrifuged for 1 hour at 32°C. This procedure was repeated with 72-hour viral supernatant. Cells expressing high
levels of GFP were subsequently sorted with the EPICS altra cell sorter (Beckman Coulter, The Netherlands).

**Cytometric Bead Array immunoassay.** Concentrations of IL1β, IL6, IL8, IL10, IL12p70 and TNF in the culture supernatants were measured simultaneously using Cytometric Bead Array (CBA) (Human Inflammation Kit, Pharmingen, Becton Dickinson, Belgium) according to the manufacturer’s instructions. Concentrations of the cytokines were obtained by comparing the mean fluorescence intensity of the samples with that of the corresponding standard curves. Flow cytometric analysis was performed using a BD FACS scan. Data acquisition and analysis was done using BD CBA software.

**IL8 ELISA.** IL8 concentrations in supernatants were assessed in dilution series by sandwich ELISA. Microtiter plates (Nunc, USA) were coated with 2 μg/ml anti-IL8 antibody (Pharmingen) in PBS for 2 hours at RT, and non-specific binding sites were blocked with 0.1% casein-PBS at 4°C overnight. Dilution series of the supernatant (1:2) were incubated for 2 hours at RT in 0.1% casein-PBS, and detected with 1:1000 biotinylated anti-IL8 antibody (Pharmingen) and 1:500 streptavidin-HRP (Pharmingen) in 0.1% casein-PBS for 1 hour. Substrate was added and the reaction was stopped after 30 minutes with 1 M H2SO4. Plates were read at 450 and 595 nm in a microplate reader (BioRad).

**Mutation screening of MT1M and MTF1.** Genomic DNA was extracted from whole blood using the Qiagen blood and cell culture DNA kit. The sequence encompassing the MTIM gene and ~600 bases upstream of the transcription start site was PCR-amplified in 25 randomly selected CD patients and 6 controls (forward primer: 5’-GAG GCC GAC CAG TGT TC-3’, reverse primer: 5’-AGT CTC TGG ATG AAA ATG TGA G-3’). The PCR product was sequenced with BigDye Terminator v3.1 Cycle Sequencing kit and analyzed on an ABI3700 analyzer (Applied Biosystems) with the former and the following primers: exon 1: 5’-CCC AGC CCA GCC CAG GAC CG-3’; exon 2: 5’-TCA CTG CCC ACT GCG TTT TTC TC-3’; exon 3: 5’-CAA GTC TAC TGC TAC CTC TC-3’.
Mutation analysis of MTF1 was performed in 95 randomly selected CD patients using flanking intronic primers for all exons and overlapping primers for 1 kb of the upstream regulatory region. PCR primers were designed using the SNPbox software (Weckx et al. 2005a). A total of 28 primer sets were used (primer sequences are available on request). PCR fragments were
sequenced with BigDye Terminator v3.1 Cycle Sequencing kit (Perkin-Elmer, Applied Biosystems), and analyzed on an ABI3730 DNA analyzer (Applied Biosystems). The sequencing trace files were analyzed for the presence of variants using novoSNP (Weckx et al. 2005b).

**CARD15 and MTF1 genotyping.** The 3 CD-associated CARD15 polymorphisms R702W (SNP8), G908R (SNP12) and 1007fs (SNP13), as well as MTF1 IVS1-128A>T, were genotyped in a cohort of 222 randomly selected CD patients and 63 controls using RFLP-PCR as previously described (Laukens et al. 2005). MTF1 IVS1-128A>T genotyping was also done by RFLP-PCR. A fragment containing the polymorphism was amplified with the following primers: forward 5’-TAA GAC TCA ATT GAT TCA GGG ACG AGA GC-3’, reverse 5’-CAG TTG TGA TGA AAA ACG TAA TGA C-3’. The PCR product was precipitated with SeeDNA (Amersham Biosciences), digested with DraI, and restriction fragments were analyzed on a 2% agarose gel. The presence of the IVS1-128T allele abolishes a DraI site, resulting in a band of 271 bp, instead of 251 bp in the presence of the IVS1-128A allele.

**Statistical analysis.** Statistical analyses were performed using SPSS software (SPSS inc., USA). Differences between groups in immunohistochemical scores and induction AUCs were calculated using the Mann-Whitney U test. Correlations between metric data were calculated using Pearson’s Rho, and those between rank data with Spearman’s Rho. Rank data for histological scores were 0-1 = 1, 1-2 = 2, 2-3 = 3 and for RNA levels <1,000 = 1, 1,000-10,000 = 2, >10,000 = 3. Odds ratios and their 95% confidence intervals (CI) were calculated with Pearson’s χ² test. Logistic regression was used to evaluate independence and interaction between parameters. Significance of differences in ELISA IL8 concentrations between time-series were calculated using a general linear model, error bars are calculated from 4 dilutions. Two-tailed probabilities were calculated, and *P*-values of less than 0.05 were considered statistically significant.

**RESULTS**

**Identification of potential candidate genes for CD.** Gene expression in normal colon biopsies of 16 CD patients and 11 healthy controls (Table 2) was analyzed on a focus microarray chip containing 6,779 expressed sequence tags. We aimed at identifying potential candidate susceptibility genes for subsequent validation and mutation screening. This process was
Reduced metallothionein expression in Crohn’s disease

performed in two steps. First, we characterized genes that were differentially expressed between CD patients and controls using an independent t-test ($P<0.01$, $N=240$). We then selected the genes that were located near a chromosomal region that has been linked to CD in at least two independent studies (Table 3). We also included the loci that were found in a study performed on a Flemish cohort (Vermeire et al. 2004a), because the patients in our study live in this area of Belgium. We employed an arbitrary distance of 5 megabases around markers that showed the highest linkage. The exact chromosomal locations of the markers and genes were explored using the genome browser at http://genome.ucsc.edu. In this way, 18 genes were selected (Table 4), and all clones of the respective genes were sequence verified and annotated correctly.

Table 2 Clinical characteristics of CD patients analyzed by microarrays

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sex</th>
<th>Familial CD</th>
<th>Age of onset</th>
<th>Behaviour</th>
<th>Disease location</th>
<th>CARD15 status</th>
<th>Operations</th>
<th>Medication</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>no</td>
<td>A1</td>
<td>B1</td>
<td>IC</td>
<td>wt</td>
<td>no</td>
<td>immunosuppressives</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>no</td>
<td>A1</td>
<td>B3</td>
<td>IC</td>
<td>ND</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>no</td>
<td>A2</td>
<td>B1</td>
<td>IC</td>
<td>mutant</td>
<td>no</td>
<td>5-ASA</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>yes</td>
<td>A1</td>
<td>B3</td>
<td>IC</td>
<td>wt</td>
<td>resection</td>
<td>corticosteroids</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>no</td>
<td>A1</td>
<td>B3</td>
<td>IC</td>
<td>mutant</td>
<td>fistula</td>
<td>5-ASA</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>no</td>
<td>A1</td>
<td>B2</td>
<td>C</td>
<td>mutant</td>
<td>resection</td>
<td>Remicade</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>no</td>
<td>A2</td>
<td>B2</td>
<td>I</td>
<td>wt</td>
<td>resection</td>
<td>5-ASA</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>no</td>
<td>A1</td>
<td>B3</td>
<td>IC</td>
<td>wt</td>
<td>resection</td>
<td>immunosuppressives</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>yes</td>
<td>A1</td>
<td>B3</td>
<td>IC</td>
<td>wt</td>
<td>fistula</td>
<td>immunosuppressives</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>yes</td>
<td>A2</td>
<td>B2</td>
<td>I</td>
<td>wt</td>
<td>resection</td>
<td>immunosuppressives</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>no</td>
<td>A1</td>
<td>B3</td>
<td>IC</td>
<td>mutant</td>
<td>fistula</td>
<td>immunosuppressives</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>yes</td>
<td>A1</td>
<td>B2</td>
<td>I</td>
<td>wt</td>
<td>no</td>
<td>5-ASA</td>
</tr>
<tr>
<td>13</td>
<td>F</td>
<td>no</td>
<td>A1</td>
<td>B3</td>
<td>C</td>
<td>wt</td>
<td>resection</td>
<td>immunosuppressives</td>
</tr>
<tr>
<td>14</td>
<td>F</td>
<td>no</td>
<td>A1</td>
<td>B1</td>
<td>IC</td>
<td>wt</td>
<td>no</td>
<td>5-ASA</td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td>no</td>
<td>A1</td>
<td>B3</td>
<td>IC</td>
<td>wt</td>
<td>fistula, resection</td>
<td>corticosteroids</td>
</tr>
<tr>
<td>16</td>
<td>F</td>
<td>no</td>
<td>A1</td>
<td>B3</td>
<td>IC</td>
<td>wt</td>
<td>fistula</td>
<td>5-ASA</td>
</tr>
</tbody>
</table>

A1: <40 years; A2: ≥ 40 years; B1: non-stricturing, non-penetrating; B2: stricturing; B3: penetrating; disease location is defined as maximal extension of inflammation during total follow-up of patients with I: ileal involvement only, C: colonic involvement only, IC: ileal and colonic involvement; mutant CARD15: carriage of at least one mutant allele for SNP8, SNP12 or SNP13 (according to Hugot et al. 2001); wt CARD15: not carrying a mutant allele for SNP8, SNP12 or SNP13; 5-ASA: 5-aminosalicylates

Metallothioneins are down-regulated in colon, ileum and whole blood of CD patients with colonic involvement.

Metallothionein RNA expression. In the microarray screen we found two closely related metallothionein (MT) transcripts, MT1F and MT1M, that were down-regulated in CD patients. Although other genes may show lower $P$-values, we chose to focus on these genes, because they are located near the IBD1 locus, which has been linked to CD in almost every linkage analysis,
Table 3 Differentially expressed genes in colon biopsies of CD patients located near a locus for CD

<table>
<thead>
<tr>
<th>Unigene symbol</th>
<th>Gene symbol</th>
<th>Gene description</th>
<th>Chromosomal location</th>
<th>Expression in CD</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hs.486246</td>
<td>PHTF1</td>
<td>putative homeodomain transcription factor 1</td>
<td>1p13</td>
<td>↑</td>
<td>0.000099</td>
</tr>
<tr>
<td>Hs.77955</td>
<td>MEF2D</td>
<td>MADs box transcription enhancer factor 2, polypeptide D</td>
<td>1q22</td>
<td>↓</td>
<td>0.0035</td>
</tr>
<tr>
<td>Hs.106674</td>
<td>BAP1</td>
<td>BRCA1 associated protein</td>
<td>3p21.31-p21.2</td>
<td>↓</td>
<td>0.0017</td>
</tr>
<tr>
<td>Hs.438691</td>
<td>GMPPB</td>
<td>GDP-mannose pyrophosphorylase B</td>
<td>3p21.31</td>
<td>↓</td>
<td>0.0013</td>
</tr>
<tr>
<td>Hs.302047</td>
<td>PLCCL3</td>
<td>phospholipase C-like 3</td>
<td>3q25.31</td>
<td>↓</td>
<td>0.0048</td>
</tr>
<tr>
<td>Hs.318567</td>
<td>NDRG1</td>
<td>N-myc downstream regulated gene 1</td>
<td>8q24.3</td>
<td>↓</td>
<td>0.00038</td>
</tr>
<tr>
<td>Hs.84072</td>
<td>TM4SF3</td>
<td>transmembrane 4 superfamily member 3</td>
<td>12q12.1</td>
<td>↑</td>
<td>0.0046</td>
</tr>
<tr>
<td>Hs.85951</td>
<td>XPORT</td>
<td>exportin, tRNA (nuclear export receptor for tRNAs)</td>
<td>12q14.1</td>
<td>↑</td>
<td>0.0021</td>
</tr>
<tr>
<td>Hs.419776</td>
<td>NAP1L1</td>
<td>nucleosome assembly protein 1-like 1</td>
<td>12q21.1</td>
<td>↓</td>
<td>0.0088</td>
</tr>
<tr>
<td>Hs.159481</td>
<td>GALGT</td>
<td>UDP-N-acetyl-alpha-D-galactosamine</td>
<td>12q13.3</td>
<td>↓</td>
<td>0.00054</td>
</tr>
<tr>
<td>Hs.438737</td>
<td>MT1F</td>
<td>metallothionein 1F</td>
<td>16q12.2</td>
<td>↓</td>
<td>0.001</td>
</tr>
<tr>
<td>Hs.188518</td>
<td>MT1M</td>
<td>metallothionein 1M</td>
<td>16q12.2</td>
<td>↓</td>
<td>0.00058</td>
</tr>
<tr>
<td>Hs.100914</td>
<td>CEP192</td>
<td>Centrosomal protein 192kDa</td>
<td>18p11.21</td>
<td>↑</td>
<td>0.000042</td>
</tr>
<tr>
<td>Hs.512640</td>
<td>PRKCSH</td>
<td>protein kinase C substrate 80K-H</td>
<td>19p13.2</td>
<td>↓</td>
<td>0.00042</td>
</tr>
<tr>
<td>Hs.134074</td>
<td>SLC35E1</td>
<td>solute carrier family 35, member E1</td>
<td>19p13.11</td>
<td>↑</td>
<td>0.006</td>
</tr>
<tr>
<td>Hs.437</td>
<td>TCF15</td>
<td>transcription factor 15 (basic helix-loop-helix)</td>
<td>20p13</td>
<td>↓</td>
<td>0.0024</td>
</tr>
<tr>
<td>Hs.102336</td>
<td>ARHGAP8</td>
<td>Rho GTPase activating protein 8</td>
<td>22q13.31</td>
<td>↓</td>
<td>0.0093</td>
</tr>
<tr>
<td>Hs.28491</td>
<td>SAT</td>
<td>spermidine/spermine N1-acetyltransferase</td>
<td>Xp22.1</td>
<td>↑</td>
<td>0.0035</td>
</tr>
</tbody>
</table>

↑: up-regulated in CD; ↓: down-regulated in CD

and which contains CARD15. Metallothioneins are involved in a variety of metal scavenging processes, and they possess both bactericidal and anti-inflammatory activities (Zangger et al. 2001; Itoh et al. 2005). In humans, at least 18 MT gene isoforms exist (MT1A-MT1X, MT2A, MT3 and MT4), but only MT1 and MT2 are inducible. We performed a detailed expression study on different MT isoforms using qPCR on an independent population of 38 CD patients and 19 controls. Because of the high sequence homology between MT isoforms, we first determined whether both MT1M and MT1F were specifically down-regulated. To this end, we designed primers for MT1M and MT1F, as well as for MT1E, MT1J and MT2A (Table 1). Primer specificities were confirmed by the presence of a single melting peak after denaturation of the amplicons, and by direct sequencing. The expression of MT1M in CD patients with colonic involvement (subgroups C and IC) was significantly less than in patients with pure ileal involvement (P=0.0244, Figure 1) and in controls (P=0.0076). Furthermore, the expression of MT1M in all biopsies tested was significantly correlated with expression of the other isoforms tested (Table 5). This indicates that at least these MT isoforms share the same regulatory machinery for basal expression in the colon. The highest expression in colon biopsies was observed for MT1E, followed by MT2A>MT1J>MT1F>MT1M.
**Table 4** Correlations of MT isoform expression in colon biopsies of CD patients and controls

<table>
<thead>
<tr>
<th>MT1E</th>
<th>MT1F</th>
<th>MT1J</th>
<th>MT1M</th>
<th>MT1A</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT1E</td>
<td>Pearson’s Rho</td>
<td>1</td>
<td>0.631</td>
<td>0.496</td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>65</td>
<td>55</td>
<td>66</td>
</tr>
<tr>
<td>MT1F</td>
<td>Pearson’s Rho</td>
<td>1</td>
<td>0.589</td>
<td>0.378</td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>55</td>
<td>62</td>
<td>15</td>
</tr>
<tr>
<td>MT1J</td>
<td>Pearson’s Rho</td>
<td>1</td>
<td>0.609</td>
<td>0.480</td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>53</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>MT1K</td>
<td>Pearson’s Rho</td>
<td>1</td>
<td>0.615</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td>&lt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT12A</td>
<td>Pearson’s Rho</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Basal transcript levels of MT1F were measured in whole blood of 15 CD patients and 13 controls, and were found to be down-regulated in CD patients with colonic involvement \((P=0.026, \text{Figure 2})\). MT1M levels were too low for reliable quantitative measurement. None of the patients and none of the controls had elevated C-reactive protein levels at the time of blood sampling. Therefore, changes in MT expression due to systemic inflammation is not likely (Vermeire et al. 2004b).

**Metallothionein protein expression.** MT protein expression was evaluated semi-quantitatively in colon and ileum biopsies of CD patients and controls using immunohistochemical staining with a mouse monoclonal anti-metallothionein antibody (cloneE9, Zymed Laboratories). This antibody however, cannot distinguish between MT isoforms, because all of them have the epitope it recognizes. Overall, the expression of MT in colon was significantly lower than in the ileum \((P=0.042)\). No conclusions could be drawn about MT expression in the colon, because the scores rarely reached 2. Compared to controls, a significant decrease in MT protein expression was found in CD patients with colonic involvement (subgroups C and IC) in ileum biopsies, again indicating a possible fundamental defect in MT expression in these patients (data not shown). A correlation was found between RNA expression as assessed by qPCR and the immunohistochemical score of protein expression \((\text{Spearman’s Rho: 0.826, } P=0.001, N=12)\).
This finding is interesting because in each individual, the biopsies used for RNA extraction and those used for staining were not obtained at the same time, again indicating a stable inherent basal expression of MT. Epithelial cells were primarily positive for MT staining. The strongest expression was found at the base of the villi, within paneth cells, and in rapidly proliferating epithelial cells in the crypts (Figure 3).

**Metallothionein induction in response to oxidative stress is higher in CD patients.** Because MT expression is decreased in the colon, ileum and whole blood of CD patients with colonic involvement, we examined whether this was due to impaired induction. We stimulated PBMC isolated from 10 CD patients (subgroups C: 2, IC: 4, I: 4) and 9 controls with known MT inducers (10 ng/ml IL6, 1000 U/ml TNF, 200 μM ZnSO4, 200 U/ml IFNγ and 2 μM dexamethasone), and measured MT1M expression by qPCR. No difference in induction was found between CD patients and controls. However, induction following exposure to 100 μM H2O2 for 10 minutes was significantly higher in CD patients, independently of the disease location ($P=0.0015$, Figure 4).

**Colonic epithelial cells defective in metallothionein expression secrete less IL8 in response to LPS or adherent-invasive E. coli.** Because we observed MT expression mainly in epithelial cells, we investigated whether a low basal expression of MT in epithelial cells has consequences upon challenge with bacteria. We used small interfering RNA (siRNA) to create HT29 colonic epithelial cells deficient in MT expression (HT29MTkd). Although the siRNA sequence was targeted against MT1B, MT1E, MT1H, MT1J and MT1M, the RNA expression of most MT isoforms was lowered approximately twofold as assessed by qPCR (Figure 5). Diminished
Reduced metallothionein expression in Crohn’s disease

Figure 2 MT1F expression in blood is down-regulated in CD patients with colonic involvement. CD colitis: subgroups IC + C (N=11), CD ileitis: subgroup I (N=4), control (N=13).

protein expression was shown by immunofluorescence using the monoclonal anti-metallothionein antibody (data not shown). Cytokine secretion in response to stimulation with adherent-invasive E. Coli (AIEC LF82) was tested with the Human inflammation kit (Becton Dickinson). We could only measure IL8 secretion after bacterial challenge with this technique. The secretion of IL8 was significantly lower in HT29MTkd cells than in normal HT29 cells (Figure 6).

Screening for mutations in the MT1M gene. We screened the complete MT1M gene (~1.4 kb) and its promoter region (~600 bases upstream of the transcriptional start site), using NM_176870 as reference sequence, in 25 CD patients and 6 controls. Two intronic polymorphisms, IVS1-366C>T and IVS2-49C>T, and one 3’ untranslated region (3’UTR) polymorphism, c.286A>G, were found. The absence of promoter or coding mutations in MT1M that would cause its down-regulation was not too surprising, since the down-regulation was not confined to MT1M. Better candidates for mutation screening are probably upstream factors regulating all MT isoforms, and so we proceeded in that direction.

A polymorphism in the first intron of the MTF1 transcription factor gene is associated with ileal disease. One of the best-characterized transcription factors regulating MT expression is MRE-binding transcription factor 1 or MTF1, an essential zinc finger protein that binds to specific DNA motifs termed metal-response elements (MRE). Furthermore, this protein is responsible for both basal and inducible expression of MT (Samson and Gedamu 1998), and the gene is located at 1p33 (IBD7), a locus that has been linked to CD in a Flemish population.
Figure 3 Metallothionein staining of representative samples of non-inflamed ileal biopsies of (A) a control and (B) a CD patient. Control ileum shows strong immunoreactivity for MT in epithelial cells, with stronger reactivity in the proliferative cells of the crypts and at the base of the villi. In CD patients, MT expression is significantly decreased. Bar=100 µm

(Vermeire et al. 2004a). We show that expression of the transcription factor MTF1 in blood of CD patients and controls correlates with the expression of MT1F (Pearson’s Rho: 0.572, \( P=0.001, N=20 \)). All 11 exons and the promoter region (~1000 bp upstream of the transcription start site) were screened for mutations in 95 CD patients, using NM_005955 as a reference sequence. Two missense mutations were found: c.198C>G (Asp63Glu) in 4 patients, and c.1253G>A (Glu385Lys) in 2 patients. The functional relevance of these mutations still needs to be investigated. Moreover, a potential polymorphism at the splice site junction between exons 8 and 9, c.1270A>G, was frequently found. Yet no alternative splice variant could be detected in cDNA from blood of patients with different genotypes (data not shown). Here, we focused on a polymorphism in the first intron of MTF1, IVS1-128A>T, because of its potential influence on gene expression (Kleinjan and van Heyningen 2005). No difference in frequency between a cohort of 222 CD patients and 63 controls was detected (data not shown). However, genotype-phenotype analysis revealed that IVS1-128A>T had considerable influence on the location of
Reduced metallothionein expression in Crohn’s disease

Figure 4 MT1M RNA induction in response to oxidative stress is higher in CD patients (N=10) than in controls (N=9), irrespective of the disease location (P=0.0015).

Disease: the IVS1-128T allele was associated with ileal disease (subgroups I and IC). In the presence of the IVS1-128T allele, 139/166 patients (84%) had ileal disease, compared to 38/56 (68%) of those with the AA genotype (OR: 2.4, CI: 1.216-4.891, P=0.011). Because we observed that CARD15 mutations were also highly correlated with ileal disease (subgroups I and IC, P<0.0001), which has also been observed by others (Ahmad et al. 2002; Cuthbert et al. 2002; Lesage et al. 2002), we investigated whether the combined presence of CARD15 and MTF1 risk genotypes had a substantially higher impact on disease location than either of them alone. The odds ratio for developing ileal disease (subgroups I and IC) in CARD15 mutation carriers is 4.2 (CI: 1.980-9.098) and in IVS-128T allele carriers 2.4 (CI: 1.216-4.891). Moreover, logistic regression revealed significant interaction between the two genetic markers: carriage of an IVS-128T allele increased the risk of ileal disease by 12% in CARD15 wild type patients (from 60 to 72%, Figure 7), and in carriers of CARD15 mutations by 23% (from 74 to 97%). This means that both genes contribute to the location of inflammation, and mutually interact.

The IVS-128A>T polymorphism is located in the first intron of MTF1, which is located within the 5’UTR. It is not contained within a CpG island. A search for transcription factor binding sites using the transcription factor database (TRANSFAC) revealed a potential binding site for GATA binding protein 4 (GATA4).

DISCUSSION

Identification of susceptibility genes, their interaction, and their relationship to specific clinical manifestations is an important step in understanding CD and developing improved clinical
management or therapy. We used an alternative, hypothesis-independent strategy to identify novel candidate genes. Transcriptome analysis of non-inflamed colon biopsies of CD patients and controls revealed 18 novel potential candidate genes localized to CD loci.

Many genome-wide linkage studies of CD have been performed (Brant and Shugart 2004b). This led to the identification of the first gene strongly associated with CD, CARD15 (Hugot et al. 2001; Ogura et al. 2001). Nevertheless, despite its strong association with CD, CARD15 is difficult to relate causally to CD. Polymorphisms show very limited penetrance, occurring in approximately 40% of patients (predominantly those with ileal disease), as well as in 15% of healthy individuals. Furthermore, linkage to chromosome 16 was still observed in CD patients not carrying common CARD15 mutations (Hampe et al. 2002; Shaw et al. 2003; van Heel et al. 2003). An alternative explanation is that polymorphisms in CARD15 are not in themselves causal, but modify the immune response in inflammatory lesions elicited by some other mechanism. An interesting theory based on evolutionary benefit was suggested by Hugot and co-workers (Hugot et al. 2003). They stated that a mutated CARD15 protein would have been beneficial during the outbreak of the plague in Europe, but it somehow represents a disadvantage in the pathogenesis of CD. The identification of CARD15 and the insights into its role in innate immunity and CD pathology (Eckmann and Karin 2005) highlight the importance of mapping susceptibility genes.

Two of the 18 potential candidate genes we identified belong to the family of closely related metallothioneins (MT). They are located in tandem within the IBD1 locus and arose by non-processed gene duplications (Karin et al. 1984). Furthermore, they are involved in protecting cells against toxic levels of metal ions, radicals and bacterial infections. Four MT classes exist in humans, but only MT1 and MT2 isoforms are inducible by cytokines, hormones, metals and

---

**Figure 5.** The RNA expression of different MT isoforms in HT29MTkd cells is down-regulated in comparison to HT29. The solid line represents a twofold change in expression. The short hairpin RNA was targeted at MT1B, MT1E, MT1H, MT1J and MT1M (black bars), but the other isoforms tested (MT1A, MT1F, MT1G, MT1X and MT2A) also show reduced expression.
Reduced metallothionein expression in Crohn’s disease

Figure 6. IL8 secretion in response to bacterial challenge is lower in MT-knockdown cells (HT29MTkd) than in HT29 cells. +AIEC: stimulated with adherent-invasive E. coli LF82 MOI 100; unst: unstimulated (P<0.05).

stress in general (Haq et al. 2003). We showed that MT was significantly down-regulated in CD patients with colonic involvement as compared to healthy controls. In addition, detailed expression analysis of different MT isoforms revealed that their expression is highly correlated at basal level. Low RNA expression levels were linked to reduced protein expression. Furthermore, we provide evidence that a low expression level of MT in CD patients with colonic involvement results from a genetic predisposition rather than from an early inflammation event. Indeed, the down-regulation was not only found in colon but also in ileum biopsies, both sampled from non-inflamed areas, as well as in whole blood samples regardless of C-reactive protein levels. Remarkably, MT RNA levels correlated with protein expression in biopsies that were not time matched, contributing to the idea that MT expression is stable and potentially inherited. Two studies have previously shown an up-regulation of MT in CD (Bruwer et al. 2001; Dooley et al. 2004), while we and others found a down-regulation (Clarkson et al. 1985; Elmes et al. 1986; Ioachim et al. 2003; Kruidenier et al. 2003). Moreover, MT quantification in CD based on radioimmunoassay (Mulder et al. 1991), microarray (Lawrance et al. 2001) and silver-saturation assay (Sturniolo et al. 1998) also reported a down-regulation.

We further demonstrated that although basal MT levels were reduced in CD patients with colonic involvement, they are still inducible by conventional stimuli in peripheral blood. Unexpectedly, MT induction with oxidative stress was higher in CD patients, independent of disease location. Relative overreaction in CD patients might reflect the fact that basal low levels of MT result in more free hydroxide radicals that can activate the MT transcription factor, MTF1, resulting in the overall increased transcription of MT (Zhang et al. 2003). Alternatively,
the higher basal MT level in controls may not permit further induction, because a certain threshold is reached. In order to clarify the potential functional consequence of reduced basal levels, we generated MT-knockdown colonic epithelial HT29 cells using small interfering RNA. We report that, in response to bacterial challenge with a CD-associated *E. coli* strain (Darfeuille-Michaud et al. 2004), reduced expression of MT in epithelial cells correlated with reduced IL8 secretion. This observation is in accordance with the recent finding that IL8 up-regulation is impaired in response to acute trauma to colon and skin of CD patients (Marks et al. 2006). IL8 is a potent chemoattractant and activator of neutrophils, which respond rapidly to different types of infections, and play an essential role in the inflammatory response. They release antimicrobial peptides as well as reactive oxygen intermediates that may cause tissue damage. We hypothesize that the intestinal epithelial cell lining of CD patients with colonic involvement expresses less MT and initially secretes less IL8 in response to bacteria. This might secure an exaggerated secondary, compensatory immune response.

It is unlikely that polymorphisms in the MT genes themselves would be responsible for their low expression, since we showed that they are down-regulated as a group. Yet, in our knockdown model for MT, we see a similar decrease in expression of MT isoforms that were not targeted with the short hairpin RNA. Thus, the possibility of a positive feedback loop for MT expression via its own transcriptional regulator cannot be ruled out (Kimura et al. 2002). We could not find coding or promoter polymorphisms in *MT1M*, but other isoforms need to be screened. Alternatively, epigenetic mechanisms such as DNA methylation might be responsible for the
Reduced metallothionein expression in Crohn’s disease

In searching for mutations in MTF1 (IBD7), a transcription factor that regulates basal expression of MT, we focused on a polymorphism within the first intron, IVS1-128A>T, because regulating regions are frequently found there. This polymorphism predisposes to inflammation at specific sites in the intestine. Disease location is a phenotypic characteristic that remains stable over time. Similar to CARD15, the association of MTF1 with ileal disease offers a new candidate disease-modifying gene, rather than a disease-predisposing gene (Gasche et al. 2003). In addition, the two genes interact to bring about ileal disease. If we regard pure colonic CD on one end of a continuum with ulcerative colitis (UC) on the other end, this polymorphism in UC patients is worthy of study.

The links of CARD15 and MTF1 with disease location might reflect the functions of these proteins at their sites of expression: CARD15 is expressed in paneth cells of the ileum, and the expression of MT in ileum is much higher than in colon. Both proteins are probably crucially involved in maintaining a low bacterial count in the ileum. A fully functional CARD15 or MTF1 might be crucial in this process. However, they might be disadvantageous in the colon, where bacterial load is high. In the colon, a weakly functional CARD15 can be beneficial, because inappropriately intense immune responses will not be elicited. Greater MT expression in the colon can also be beneficial because it can lead to efficient clearance of mucosal infection.

The association of the IVS1-128T allele in MTF1 with low MT expression needs investigation. However, to fully understand the genetic contribution of IVS-128A>T in CD, the haplotype structure surrounding this polymorphism needs to be determined. The mutation causing altered transcription of MTF1 could be located elsewhere within the same haplotype block. MTF1 regulates basal expression of MT (Samson and Gedamu 1998). Therefore, a difference in MTF1 expression will be directly linked to altered MT expression. Indeed, we showed a correlation between MT1F and MTF1 levels in peripheral blood. The IVS-128T allele is located within a potential binding site for GATA4, a transcription factor that is expressed in the ileal enterocytes in the small intestine, but not in the colon (Boudreau et al. 2002). Therefore, a primary goal in this context is to study GATA4 binding to the MTF1 gene.

Loss of MTF1 in mouse embryonic fibroblasts results in enhanced collagen deposition, which is an important complication of CD (Haroon et al. 2004). In these fibroblasts, transforming growth factor-beta is activated. This protein has potent anti-inflammatory properties, but at the same time it drives the process of fibrosis in the deeper layers of the gut (Van Assche et al. 2004). Nonetheless, we could not find an association between the IVS-128A>T polymorphism and stricturing disease, indicating that additional risk factors influence this particular phenotype.
Interestingly, though, the IVS-128T allele in MTF1 is associated with CD of the ileum, which has a higher likelihood of stricturing than does colonic CD. We used microarray screenings in unaffected tissues for human genetic studies. A similar study was performed by Lawrance and colleagues, with the exception that they used moderately inflamed resected colonic tissue (Lawrance et al. 2001). We used non-inflamed tissue to target basal differences in gene expression due to genetic variation, and not due to inflammation-related events. Significant evidence was found for the genetic transmission of variation in gene expression (Lo et al. 2003; Pastinen et al. 2004). The variation in expression level is highest among unrelated individuals, and smallest between monozygotic twins, indicating that germ-line differences contribute to variation in gene expression (Cheung et al. 2003; Correa and Cheung 2004).

In summary, screening unaffected colon biopsies by microarrays proved to be useful in the identification of new candidate genes for CD. A cluster of MT genes located at IBD1 is consistently down-regulated in CD patients with colonic involvement. Subsequently, we identified a new disease-modifying gene, MTF1, that is associated with ileal disease, and together with CARD15 serves as a good predictor of disease location in CD patients.

AKNOWLEDGEMENTS

We acknowledge the contribution of the VIB Genetic Service Facility (http://www.vibgeneticservicefacility.be) to the genetic analyses, and the Protein expression and Purification core facility of VIB for providing recombinant TNF.

REFERENCES


Reduced metallothionein expression in Crohn's disease

10(3): 300-11.


Reduced metallothionein expression in Crohn’s disease


Chapter 3


Chapter 4

HUMAN METALLOTHIONEIN EXPRESSION UNDER NORMAL AND PATHOLOGICAL CONDITIONS: MECHANISMS OF GENE-REGULATION

Debby Laukens¹, Anouk Waeytens², Pieter De Bleser¹, Claude Cuvelier², Martine De Vos³ and Pieter Rottiers¹

¹Department of Molecular Biomedical Research, Ghent University and Flanders Interuniversity Institute for Biotechnology (VIB), Gent, Belgium
²Department of Pathology, Ghent University Hospital, Gent, Belgium
³Department of Gastroenterology, Ghent University Hospital, Gent, Belgium

ABSTRACT

Metallothioneins (MT) are ubiquitous metal-binding proteins that are highly conserved throughout evolution. Although the exact physiological function is not completely understood, it is clear that they are involved in a variety of processes including metal detoxification, free radical scavenging, metal homeostasis and cell proliferation. The human MT family consists of at least 18 different isoforms, containing pseudogenes as well as functional proteins. They can be induced by a wide variety of substances, e.g. metals, cytokines and hormones. In addition, different cell types express discrete MT isoforms, reflecting the specifically adapted functions of MT isoforms, and hence a divergence in their regulation. Aberrant expression of MT has been described in a number of apparently diverse diseases, including inflammatory bowel disease, cancer, Alzheimer’s disease, amyotrophic lateral sclerosis and Menkes disease. Therefore, a thorough understanding of the regulation of MT expression is important. To date, the regulation of transcription of these genes has primarily been studied in mice. Unfortunately, the situation in mice is somehow less complicated, since only 4 isoforms are expressed. Nevertheless, the high homology between mouse and human MTs allows us to evaluate regulatory regions in their respective promoters. Here, we review the aberrant expression of human MT in disease, and the mechanisms that regulate MT expression.
Chapter 4

INTRODUCTION

Metallothioneins (MT) are a family of small, highly conserved proteins with the specific capacity to bind metal ions. The MT protein was first purified from the equine renal cortex in 1960 (Kagi and Valee 1960). Since, they were described in a wide variety of species, including vertebrates, invertebrates, plants, fungi and some prokaryotes. A great deal of sequence and structural homology exists between MT proteins in different species, underlining its important biological role. Mammalian MT proteins typically consist of 61 to 68 amino acids, with a high content of polar, highly catalytic cysteine residues. These cysteins are strictly conserved and arranged in motifs that form the framework of two distinct metal-binding domains, linked by a short peptide (Figure 1). Initially, the biological function of MT was centralized to their unique metal-binding capacity. In normal conditions, excessive concentrations of essential and nonessential metal ions like cadmium, mercury, and lead can be toxic. Most organisms use a redundant array of cellular mechanisms to limit toxicity of metal ions (Dameron and Harrison 1998), one of which is sequestration by MT. They thus provide homeostasis of metal ions and protect cells from acute heavy metal toxicity. Through participating in zinc metabolism, they regulate the activity of fast exchanging metalloproteins, such as NFκB (Sakurai et al. 1999; Kim et al. 2003) and the tumor suppressor gene p53 (Ostrakhovitch et al. 2006). It is now clear that the function of MT is not restricted to this metal-binding activity. They are rapidly induced during specific phases of the cell cycle and by diverse stimuli, such as metals, hormones and cytokines, thus they participate in cell cycle and diverse protective functions. Metallothioneins are capable of scavenging free radicals and thus play a role in protection of tissues against various forms of oxidative injury, including radiation, lipid peroxidation, oxidative stress conditions of hyperoxia. In addition, extracellular MT has been shown. Elevated serum and urine MT was found in cadmium-exposed individuals (Falck et al. 1983). Moreover, MT in the extracellular environment may act as a “danger-signal” promoting movement of leukocytes to the site of inflammation (Yin et al. 2005). A significant role for a strictly controlled regulation of MT in both growth of cells and in their response towards several stimuli is obvious. It is therefore plausible that defects in the one of these restricted mechanisms can lead to pathological situations, like for example tumor growth. Good knowledge of their regulation is thus crucial in understanding their role in pathogenesis.

THE HUMAN METALLOTHIONEIN FAMILY

The classification of metallothioneins has been revised in 1999, and is based on evolutionary
Figure 1 Homology model of human metallothionein 1A. Cysteine residues are shown as yellow sticks, metal ions are shown in red (Cd$^{2+}$), green (Zn$^{2+}$) and orange (Na$^+$). The model was constructed using the Swiss Model web server (Schwede et al. 2003) with the crystal structure of the rat metallothionein 1A protein (PDB ID 4MT2) as template structure.

data (Binz 1999). Rodents have four MT isoforms, MT1 to MT4, while all primates examined so far contain multiple copies of the MT1 isoform. The situation is most complex in humans: a total of 18 MT isoforms and 5 MT-like genes have been cloned so far, many of which only differ in distinct amino acids (Table 1). At least five of the isoforms are non-processed pseudogenes. Seventeen out of 18 isoforms cluster together on chromosome 16q13 (Figure 2) (Karin et al. 1984). Apparently, chromosome 16 is one of the most enriched chromosomes for intrachromosomal duplications as compared to the human genomic average (Martin et al. 2004), and includes for example the cadherin gene cluster. It is not clear whether some of the MT pseudogenes are functional: MT1J and MT1M contain a promoter, are transcribed, but contain a premature stop codon.

The various MT isoforms differ mainly in their expression pattern: MT3 an MT4 are constitutively expressed in specific cell types, while MT1 and MT2 are highly inducible and ubiquitously expressed (Table 1).

By comparing the coding sequences of rodent and primate MT genes, it was shown that human MT1 isoforms show less divergence from human MT2 than from mouse MT1, suggesting that these two isoforms arose after the emergence of primates (Schmidt et al. 1985). Similarly, MT3 and MT4 probably diverged prior to the primate/rodent divergence (Figure 3). In addition, MT1 and MT2 isoforms are clearly separated from the MT3 and MT4 clusters, probably reflecting their unique functions in the organism. Moreover, this means that MT1 and MT2 proteins might have a very similar function in mice and in humans.
**Table 1** Annotated human metallothionein genes and their characteristics

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Chromosomal location</th>
<th>Aliases</th>
<th>Protein length</th>
<th>Refseq status</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT4</td>
<td>16q13</td>
<td></td>
<td>62</td>
<td>provisional</td>
<td>squamous epithelium</td>
</tr>
<tr>
<td>MT3</td>
<td>16q13</td>
<td>GIF</td>
<td>68</td>
<td>validated</td>
<td>brain, kidney, reproductive system</td>
</tr>
<tr>
<td>MT2A</td>
<td>16q13</td>
<td>MT2</td>
<td>61</td>
<td>provisional</td>
<td>ubiquitous</td>
</tr>
<tr>
<td>MT1A</td>
<td>16q13</td>
<td>MT1S</td>
<td>61</td>
<td>validated</td>
<td>ubiquitous</td>
</tr>
<tr>
<td>MT1B</td>
<td>16q13</td>
<td>MT1Q</td>
<td>61</td>
<td>provisional</td>
<td>ubiquitous</td>
</tr>
<tr>
<td>MT1E</td>
<td>16q13</td>
<td></td>
<td>61</td>
<td>validated</td>
<td>ubiquitous</td>
</tr>
<tr>
<td>MT1F</td>
<td>16q13</td>
<td></td>
<td>61</td>
<td>provisional</td>
<td>ubiquitous</td>
</tr>
<tr>
<td>MT1G</td>
<td>16q13</td>
<td></td>
<td>61</td>
<td>provisional</td>
<td>ubiquitous</td>
</tr>
<tr>
<td>MT1H</td>
<td>16q13</td>
<td>MT-0</td>
<td>61</td>
<td>provisional</td>
<td>ubiquitous</td>
</tr>
<tr>
<td>MT1J</td>
<td>16q13</td>
<td>MT1NP</td>
<td>40</td>
<td>provisional</td>
<td>?</td>
</tr>
<tr>
<td>MT1M</td>
<td>16q13</td>
<td>MT1K</td>
<td>61</td>
<td>provisional</td>
<td>ubiquitous</td>
</tr>
<tr>
<td>MTM</td>
<td>16q13</td>
<td></td>
<td>49</td>
<td>provisional</td>
<td>?</td>
</tr>
<tr>
<td>MT1X</td>
<td>16q13</td>
<td></td>
<td>61</td>
<td>provisional</td>
<td>ubiquitous</td>
</tr>
<tr>
<td>MT pseuogenes:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT1I</td>
<td>16q13?</td>
<td></td>
<td>-</td>
<td>withdrawn</td>
<td>-</td>
</tr>
<tr>
<td>MT1L</td>
<td>16q13</td>
<td>MT1R</td>
<td>-</td>
<td>provisional</td>
<td>-</td>
</tr>
<tr>
<td>MT1P</td>
<td>16q13</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MT1D</td>
<td>16q13</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MT2P</td>
<td>4q13</td>
<td></td>
<td>-</td>
<td>provisional</td>
<td>-</td>
</tr>
<tr>
<td>MT-like:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTL1</td>
<td>1p</td>
<td></td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>MTL2</td>
<td>1p22</td>
<td></td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>MTL3</td>
<td>18</td>
<td></td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>MTL4</td>
<td>20</td>
<td></td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>MTL5</td>
<td>11q13</td>
<td>TESMIN</td>
<td>178</td>
<td>reviewed</td>
<td>testis</td>
</tr>
</tbody>
</table>

**METALLOTHIONEIN IN DISEASE**

Metallothionein expression has been studied in various pathological conditions. Unfortunately, contradictory results complicate the interpretation of the involvement of these proteins in diseases. Expression data obtained by immunohistochemistry, semi-quantitative PCR and more

**Figure 2** Genomic organization of the human metallothionein family on chromosome 16. Pseudogenes are represented in grey.
recent techniques such as quantitative PCR (qPCR) are not always comparable. The early studies relied on immunohistochemistry using antibodies, which cannot distinguish between MT isoforms, because they share the antigenic epitope. Furthermore, in normal tissues, MT expression is usually undetectable by immunohistochemistry, except in myoepithelial (van den Oord and De Ley 1994), renal (Mitropoulos et al. 2005), intestinal (Laukens et al. 2006) and thyroid epithelial cells, pancreas (Tomita and Matsubara 2000) and fetal liver (Fuller et al. 1990). On the other hand, strong staining was found in pathological tissues, especially in many tumors (Theocharis et al. 2004). Here, we summarize data of deviant MT expression in several diseases.

![Figure 3 Phylogenetic tree of human and mouse metallothionein protein isoforms. Metallothionein 3 and 4 probably arose prior to the divergence of primates and rodents, while MT1 and MT2 genes arose later. Mouse MT genes are designated as mMT1, mMT2, mMT3 and mMT4.](image)

**Metallothionein and inflammatory bowel disease**

Inflammatory bowel disease (IBD), comprises both Crohn’s disease (CD) and ulcerative colitis (UC). They are chronic inflammatory diseases of the intestine. In CD, the complete bowel wall is affected, while in UC, only the superficial layers become inflamed. In the normal intestine, we found that MT RNA and protein expression is highest in the ileum as compared to other compartments of the gut (Laukens et al. 2006). Metallothionein is localized in the enterocytes, predominantly at the base of crypts, probably due to the high proliferative state of these cells. The immunopositivity decreases towards the top of the villi. Focal staining is often observed in the intestine, as a result of the uniform staining in distinct single crypts. It was suggested that this might be the consequence of somatic mutations in stem cells, leading to strong clonal expression...
of MT in the entire crypt (Jasani et al. 1998). Aberrant protein expression was found in tissue samples originating from IBD patients. Down-regulation as well as up-regulation has been reported. Two papers reported an up-regulation (Bruwer et al. 2001; Dooley et al. 2004), while we and others found a down-regulation of MT in IBD (Clarkson et al. 1985; Elmes et al. 1986; Ioachim et al. 2003; Kruidenier et al. 2003; Laukens et al. 2006). Moreover, MT quantification in IBD based on radioimmunoassay (Mulder et al. 1991), microarray (Lawrance et al. 2001) and silver-saturation assay (Sturniolo et al. 1998) also reported a down-regulation. Using qPCR, we were able to show a decreased expression of a large proportion of MT1 isoforms and MT2 at basal level in colon biopsies, ileum biopsies and whole blood samples of CD patients that have colonic involvement (Laukens et al. 2006). This lowered mRNA expression level correlated with protein expression, even though biopsies were not acquired at the same time. These results suggested that the deficient MT expression in CD patients with colonic disease is genetically determined, and not the result of an early inflammation event. Interestingly, the MT gene cluster is located in *IBD1*, a locus that was significantly associated to CD in many genome scans (Mathew and Lewis 2004). In murine IBD models, however, MT does not appear to influence the development or progression of intestinal pathology in the DSS (Oz et al. 2005).

Zinc is often decreased in IBD patients. This element plays an important role in the prevention of free radical formation and in protection of biological structures from damage (Stefanidou et al. 2006). It was shown that dietary zinc causes an MT increase in all gut regions in rats (Tran et al. 1999; Szczurek et al. 2001). On the other hand, zinc supplementation in IBD patients did not change concentrations of MT in plasma and erythrocytes (Mulder et al. 1994), and the MT concentration in both inflamed and non-inflamed intestinal mucosa was only slightly higher. Histological inflammation scores of intestinal biopsies, plasma albumin levels, and the disease activity index of the patients did not change during the study. However, in this study, only inactive to moderately active patients were included.

A decreased MT level in CD patients could indicate a hampered maintenance of free radicals. Indeed, human monocytic cells that were transfected with an MT-antisense vector, produce more hydrogen peroxide than control THP-1 cells in the absence of a stimulus (Leibbrandt et al. 1994). The exact role of MT in CD pathogenesis has to be established in future studies.

**Metallothionein and cancer**

Immunocytochemically detectable MT overexpression was described in a variety of human tumors, and has been extensively reviewed recently (Theocharis et al. 2004). As such, we will not focus on this issue here.
The regulation of human metallothionein

The expression of MT is not universal to all human tumors, but may depend on their differentiation status and proliferative index. Metallothionein overexpression is associated with resistance to anticancer drugs and is combined with a poor prognosis. However, its use as a marker of tumour differentiation, cell proliferation and prognosis predictor remains unclear. On the other hand, gastric carcinomas and colorectal adenomas are apparently accompanied by a decreased expression of MT, however, those with a relatively high level seem to have an increased malignant potential (Janssen et al. 2000).

Cell-type specific differential regulation of human MT genes was found in different cancer cell lines, correlating with DNA methylation and chromatin structure (see later) (Jahroudi et al. 1990). Tumor cell-lines arising from paraxial mesoderm and endoderm have MT2A and MT1E genes in fully inducible form and the MT1F in the refractory state (Schmidt et al. 1985). On the other hand, tumours originating from ectoderm, intermediate and lateral mesoderm exhibit MT2A and MT1F genes in inducible form and the MT1E gene in a refractory form.

Metallothionein, Alzheimer’s disease and amyotrophic lateral sclerosis

Metallothionein 3 was first cloned as growth inhibitory factor (GIF), which showed a decreased RNA expression in Alzheimer’s disease (AD). Metallothionein 3 suppresses the neurotrophic activity present in the normal human brain (Tsuji et al. 1992b). The down-regulation of MT3 in AD has been confirmed by two recent studies. In the first study, MT3 expression was determined in a large number of AD cases by qPCR as well as by immunohistochemistry and Western blotting (Yu et al. 2001). In the second study, DNA microarrays were used to compare RNA levels from control and AD hippocampal regions and found, amongst others, MT3 down-regulation (Colangelo et al. 2002). However, Erickson and colleagues (Erickson et al. 1994) disputed that neuronal changes in AD are related to a decrease in MT3, since they could not find a significant down-regulation in neither RNA nor protein expression in their AD population. There is more consensus on the overexpression of MT1 and MT2 in the astrocytes from AD as well as other neurological disorders (Duguid et al. 1989; Adlard et al. 1998; Zambenedetti et al. 1998; Chung et al. 2004). In the brain, astrocytes are the main source of MT1 and MT2, although other cell types, such as choroid plexus epithelia, endothelium and meningeal cells may also express these isoforms (Penkowa 2006). In neurodegenerative diseases such as AD, astrocytes become abundant and activated in the affected areas. While in other organs, the main function of MT is related to zinc metabolism and protection against heavy metal and/or oxidative damage, the key role of MTs in the brain seems to be a protection in the cellular response to neuronal injury. It was suggested that the specific increase in MTs was associated with the initial
stages of the disease process (Adlard et al. 1998). The precise mechanisms downstream of MT have not been fully established, but convincing data showed that they are essential in dealing with neuropathology and for brain recovery in AD as well as other brain pathologies. MTs might even be used as therapeutic and/or preventive drugs for a range of brain disorders (Penkowa 2006).

Amyotrophic lateral sclerosis (ALS) is a progressive, invariably fatal neurological disease due to degeneration of the nerve cells responsible for controlling voluntary muscles. In ALS, motor neurons in the brain stem, spinal cord and motor cortex degenerate or die, ceasing to send messages to muscles. Consequently, the muscles gradually weaken, waste away, and twitch. Fifteen to 20% of cases of familial amyotrophic lateral sclerosis are associated with mutations in the superoxide dismutase 1 gene (SOD1). Elevated levels of MT have been found in spinal cord, kidney and liver of patients with ALS (Sillevis Smitt et al. 1992a; Sillevis Smitt et al. 1992b), but not in serum (Sillevis Smitt et al. 1994). A detailed study on different MT isoform expression revealed no evidence for either the induction of a specific MT repertoire, or for the inability of glia to express any MT gene (Blaauwgeers et al. 1996). Probably, the enhanced expression of MT in ALS reflects an early protective function. This was also concluded from a study on MT expression in mice carrying the SOD1 mutation (Gong and Elliott 2000). These mice were backcrossed with MT-knockout mice. The offspring reached the onset of clinical signs significantly earlier in response to the reduction of protein expression. These results indicated that the copper-mediated free radical generation derived from mutant SOD1 might be related to the degeneration of motor neurons in ALS and that MT might play a protective role against the expression of the disease (Nagano et al. 2001).

Recently, MT3 was screened for mutations in 20 patients with ALS, but no functionally relevant polymorphism could be associated with the disease (Morahan et al. 2005).

**Metallothionein, Menkes disease and Wilson disease**

Menkes disease is an X-linked, recessive disorder of the copper metabolism that occurs in less than 1 in 200,000 live births. The condition is characterized by early retardation in growth, peculiar hair, focal cerebral and cerebellar degeneration, skeletal abnormalities, and patient mortality in early childhood (Bankier 1995). Three independent research groups cloned the Menkes gene, a copper transporting ATPase, ATP7A, to the long arm of the X chromosome (Chelly et al. 1993; Mercer et al. 1993; Vulpe et al. 1993). A spectrum of mutations adversely affecting protein expression have been observed in severely affected Menkes patients (Kodama and Murata 1999). The genetic defect in Menkes syndrome leads to a progressive copper
The regulation of human metallothionein
deficiency and copper-dependent enzymes fail in most tissues.
Using cultured fibroblasts, it was shown that low extracellular copper concentrations induce
synthesis of MT in Menkes’ cells but not in normal cells (Hamer 1987). Therefore, in the early
studies on tissue abnormalities in Menkes disease, MT were believed to be involved in
pathogenesis (Garnica et al. 1978; Schmidt et al. 1984). Now, it is clear that a defect in efflux
and consequently intracellular accumulation of copper, due to mutations in \textit{ATP7A}, is
responsible for this up-regulation in Menkes’ cells. As a result of accumulation of copper in the
cell MTs are up-regulated, but they are as such not causative for pathogenesis in Menkes disease.
In affected cells, copper accumulates bound to MT in the cytosol, while its transport to the
organelles, as well as copper efflux, is disturbed. The low activity of metalloenzymes is believed
to contribute significantly to the pathogenesis of this condition.
Wilson disease is an autosomal recessive disorder characterized by dramatic build-up of
intracellular hepatic copper with subsequent hepatic and neurologic abnormalities. Copper
toxicity occurs when the liver is overloaded and non-ceruloplasmin-bound copper is released
into the bloodstream, from where it can diffuse into the brain. Today, the treatment of Wilson
disease is no longer aimed at ‘decoppering’, the removal of accumulated copper, but at the
normalization of the free copper concentration in blood, to reverse the copper poisoning.
Therefore, new therapy is aimed at administration of zinc to these patients, because this
increases MT expression and sequesters the excess of copper in the blood (Hoogenraad 2006).

HUMAN METALLOTHIONEIN GENE REGULATION

Because MT expression is involved in a number of pathological conditions, the transcriptional
control of MT has become a major topic. Metallothioneins are induced by a wide variety of
physiological and chemical agents like cytokines, metals, hormones, and stress in general
(Borghesi and Lynes 1996). Furthermore, they are transiently induced after tissue injury caused
by e.g. inflammation or irradiation (Manuel et al. 1992). It is generally accepted that \textit{MT1} and
\textit{MT2} genes are inducible, while \textit{MT3} and \textit{MT4} are constitutively expressed. Metallothionein 1
isoforms have a restricted and transient role, perhaps in stress or during infection and the cell
cycle. On the other hand, MT2A is ubiquitously expressed, and plays a general role in cellular
physiology, possibly in zinc metabolism. Metallothionein 3 is predominantly expressed in the
central nervous system (Tsuji et al. 1992a; Blaauwgeers et al. 1996), but also in the kidney
(Hoey et al. 1997), prostate (Dutta et al. 2002), retina (Tate et al. 2002), salivary glands (Irie et
al. 2004) and reproductive system (Moffatt and Seguin 1998). The mouse MT4 is expressed in
stratified squamous epithelium of the tongue (Quaife et al. 1994), but there are so far no data on human MT4 expression.

Most of the functional studies on MT transcription were performed in mice. However, in mice, MT1 and MT2 isoforms are co-ordinately regulated (Searle et al. 1984), while the many human MT isoforms are regulated in a cell-type specific manner (Schmidt and Hamer 1986; Varshney et al. 1986; Laukens et al. 2006). This probably resulted from the adaptation of the MT isoforms to more specific functions throughout evolution. Phylogenetic tree analysis of the promoter region of mouse and human MT isoforms shows that MT2, MT3 and MT4 promoters are highly homologue between mouse and human, which might reflect their more strictly regulated expression pattern. Human MT1 isoforms are dispersed into two groups. Interestingly, these two clusters correlate to the physical location of the genes on chromosome 16 (Figure 2). On the other hand, the mouse MT1 promoter region is relatively unrelated to the human MT1 group (Figure 4). It is thus plausible that regulatory regions in the mouse and human promoter have adapted differently. Therefore, care must be taken when evaluating MT expression and maybe even function in the mouse, for solving questions related to human circumstances. Nevertheless, if common transcription factor binding sites are found in human and mouse promoters, we can extrapolate the experimental data on mouse MT1 regulation to the human situation.

We performed a comparative study on the transcriptional control of MT in mice and humans, using in silico data of promoter transcription factor binding sites. Based on these common regulatory sites, we review the MT regulatory mechanisms.

**In silico analysis of human and mouse MT1 promoters**

Although marked sequence conservation exists, the RNA level for each MT isoform is unique. This is possibly due to inherent differences in promoter regulatory sequences. Promoters are organized with a variety of elements that contribute to promoter function. The elements found in any promoter differ in number, location and orientation.

A difference distance matrix approach (De Bleser et al. 2006) was used to identify a set of transcription binding sites that are specifically highly present in following MT isoforms: MT1A, MT1B, MT1E, MT1F, MT1G, MT1E, MT1J, MT1M and MT1X (Table 2). Many of these binding sites and their transcription factors have been described in MT regulation, however, we found a number of unexplored regulatory regions. We next searched for these transcription factor binding sites in the mouse MT1 promoter, and for well-known mouse binding sites (USF, MLTF) in the promoters of human MTs. Below, we describe the involvement of these transcriptional regulators in more detail.
Figure 4 Phylogenetic tree analysis of the promoter regions of human and mouse MT isoforms. Promoter regions are defined from nucleotides -800 to -1 relative to the transcriptional start site.

Table 2 Transcription factor binding sites in 800 bp of the promoter of human and mouse metallothionein 1 isoforms

<table>
<thead>
<tr>
<th>Transcription factor</th>
<th>MT1A</th>
<th>MT1B</th>
<th>MT1E</th>
<th>MT1F</th>
<th>MT1G</th>
<th>MT1H</th>
<th>MT1J</th>
<th>MT1M</th>
<th>MT1X</th>
<th>mMT1</th>
</tr>
</thead>
<tbody>
<tr>
<td>basal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBP</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>TFII-I</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Sp1</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>AP2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>USF/Nrf2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>induction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ChCh</td>
<td>9</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>HELIOS</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>E2F</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Sp1</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Egr 1</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>GR</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>STAT</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>MTF1</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>RAR-α</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>repression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETF</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>9</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>NFI</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

The number of hits as found by Match™ (TRANSFAC PRO 8.4) using a core matrix match of 100% and a matrix match of 85%. Exceptions are IL6RE and MLTF for which consensus sequences were used in combination with Patch™ (TRANSFAC PRO 8.4). The Match procedure introduces false positives; therefore, the number of hits listed is an overestimation of the real hits. The MT isoforms with TBP consensus sequences located close to the transcription start site are denoted with a plus (+). Those that do not contain a TBP close to the transcription start site are denoted with a minus (-). mMT1= mouse MT1
**Basal expression**

RNA polymerase II is responsible for transcribing genes coding for messenger RNA (mRNA). The first step in transcription is the binding of the TFIID complex to a region upstream of a sequence called the TATA box. The TFIID complex consists of the TATA-binding protein (TBP) and TBP-associated factors (TAF). The location of the TATA box with respect to the start point is relatively fixed, usually located ~25 bp upstream of the transcriptional start site. Therefore, TATA box consensus sequences located further upstream are probably not functional. Metallothionein 1E, MT1F, MT1H, MT1J and MT1M contain TATA-less promoters. This could partly explain the observation that the MT1G promoter is five times more active than the MT1F promoter in transfection studies (Gedamu *et al.* 1987; Shworak *et al.* 1993). GC boxes, common promoter components involved in basal transcription, are frequently found in MT promoters, often clustered together close to the start site. It has been proposed that the Sp1 factor binds to multiple GC boxes, resulting in an interaction of GC box-bound Sp1 factors with each other to synergistically stimulate transcription. Furthermore, basal level enhancer sequences, binding activator protein-2 (AP2), are found in some human MTs, but not in the mouse MT1. TFII-I binds specifically to initiator elements (Inr), supporting basal transcription. An E box, binding the upstream stimulatory factor (USF) was only found in the mouse MT1 and the human MT1G. It was shown that TFII-I also binds to upstream E box, and that TFII-I and USF interact cooperatively at both Inr and E box sites (Roy *et al.* 1991). The USF sequence overlaps with an antioxidant response element (ARE). These AREs are usually found in genes responsive to free radicals, through interaction of ARE with NF-E2-related factor 2 (Nrf2). However, this sequence is only found in the mouse MT1 and human MT1G, suggesting that they might be more responsive to free radical exposure. Nevertheless, response to oxidative stress can also be mediated by metal responsive elements (MRE) in the promoter (Dalton *et al.* 1994). MREs are recognized by the MRE-binding transcription factor 1 (MTF1), and are classically required for metal induction of MTs (see below). Nevertheless, they also participate in basal transcription. Indeed, the basal expression of MT is highly correlated to the activity of MTF1 (Ghoshal and Jacob 2001). Furthermore, we have recently shown that basal MT levels in whole blood vary considerably between individuals, but they correlate well with MTF1 expression (Laukens *et al.* 2006).

**Inducible expression**

*Metal induction.* Metallothionein 1 and MT2 isoforms are highly inducible by many metal ions, including zinc, cadmium, bismuth, mercury, copper, nickel and cobalt. The concentration of
metal ions to induce MTs depends on the type of ion and MT isoform. Metallothioneins are capable of binding most of these elements, however, they do not bind nickel and cobalt. The transcriptional regulation responsible for this metal induced expression, is controlled by MREs, present in the promoter of MT as multiple, non-identical copies. The mouse and human transcription factor that binds to these MRE elements is the MRE-binding transcription factor 1 or MTF1. When this factor was first cloned (Brugnera et al. 1994), it was shown that the human MTF1 was more effective than the mouse equivalent. MTF1 is absolutely necessary for both basal and metal inducible MT expression (Heuchel et al. 1994). Moreover, although MTF1 is activated by a number of metals, it absolutely requires zinc for its activity. It was hypothesized that inducing non-zinc metals can displace zinc from its storage proteins, resulting in a pool of free zinc available for activation of MTF1 (Jacob et al. 1998). These storage proteins could be MT itself, leading to a complex feedback interaction between MTF1 and MTs.

Interestingly, treatment of cells with cadmium increases MT expression, although it does not influence the DNA binding activity of MTF1 to the MRE. This suggests that additional mechanisms play a role in metal induction. For example, cadmium induces oxidative stress, which could activate binding of USF to the E box. Alternatively, as stated before, cadmium might replace the intracellular zinc in storage proteins, which results in more free zinc to activate MTF1.

In addition, posttranslational modification of MTF1 has been shown. Phosphorylation of MTF1 plays a critical role in its activation by zinc and cadmium (Saydam et al. 2002). Several phosphorylation sites are present throughout the complete MTF1 protein. This was thought to be mediated through a complex pathway involving protein kinase C, tyrosine kinase, and casein kinase II.

**Stress and inflammation mediated induction.** Similar to acute phase proteins, MT is induced by inflammation, bacterial infection and stress. Stress in general often results in the synthesis of glucocorticoid hormones, resulting in the suppression of inflammation and an increase in blood sugar levels. These steroid hormones are synthesized by the adrenal gland, and can enter the cell by simple diffusion. Within the cell, it binds to its receptor, the glucocorticoid receptor (GR), which in turn gets activated and translocates to the nucleus. There, it has high affinity for a consensus sequence called the glucocorticoid response element (GRE), and activates transcription from MTs and other GRE containing genes. Glucocorticoid hormones have been known for a long time as inducers of MT (Karin and Herschman 1979; Jacob et al. 1999). The human MT2 gene contains one GRE, while the mouse MT1 and MT2 genes contain two tandem
copies ~1kb upstream of the MT2 gene and ~7 kb of the MT1 gene (Kelly et al. 1997). In humans, MT2A is significantly more inducible by glucocorticoids as compared to MT1 isoforms (Schmidt and Hamer 1986). It was shown that dexamethasone, a synthetic glucocorticoid agonist, appears to have no significant effect on the expression of MT1F (Varshney et al. 1986). This can be explained by the fact that there is no GRE sequence in the promoter of MT1F.

During acute inflammation, such as after tissue damage or during infection, macrophages are recruited and activated to the site of inflammation. They secrete pro-inflammatory cytokines such as IL6, TNF, IL1α and IL1β. These cytokines are able to induce MT (De et al. 1990). Fast MT up-regulation after challenge with IL1 is probably mediated through glucocorticoids (Coto et al. 1992). Interleukin 6 is one of the most potent inducers of MT (Schroeder and Cousins 1990). Metallothioneins have IL6 response elements (IL6RE) in their promoter. These DNA stretches bind STAT transcription factors. A synergistic effect was demonstrated between IL6 and glucocorticoid in MT induction (Kasutani et al. 1998), which is possibly mediated by the close proximity of the GRE and the IL6RE in the MT promoters.

Interestingly, it was reported that MTs inhibit the release of pro-inflammatory cytokines (Kanekiyo et al. 2002; Inoue et al. 2006). This could be explained by the regulatory role of MT in NFκB activation (Sakurai et al. 1999).

Cell cycle. Metallothioneins appear to play a major role during the cell cycle. A tenfold rise of MT synthesis was described in exponentially growing human hepatocytes (Nagel and Vallee 1995; Studer et al. 1997), and peaks of MT expression were found in late G1 and G1/S transition in HT29 epithelial cells (Nagel and Vallee 1995). Similarly, in placental tissue, positive immunostaining for MT was found only in trophoblast and proliferating cells (Haerslev et al. 1995). We have recently described that MT expression in the intestine was most apparent in the rapidly proliferating cells of the crypts (Laukens et al. 2006). Moreover, it was suggested that MT transcription is altered by the differentiation process. Indeed, it was shown that the differentiation of teratocarcinoma cells using retinoic acid is associated with a rise in MT expression. Retinoic acid receptors (RAR) are nuclear receptors related to the steroid and thyroid hormone receptors, a family of proteins that functions as ligand-dependent transcription factors. Retinoic acid is a regulator of differentiation at various stages of vertebrate embryogenesis. In accordance, multiple RAR receptor binding sites are found in human and mouse MT1.

Although MT expression is generally cytosolic, nuclear translocation has been observed at G0/G1 to early S-phase (Cherian and Apostolova 2000). This nuclear and cytoplasmic localization of MT was also observed in several tumours, especially in regions of high
proliferation. Moreover, antisense down-regulation of MT1 in endothelial cells resulted in the cell cycle arrest at the G1 phase (Miyashita and Sato 2005), and cell growth was inhibited in MT1 antisense tumor cells (Takeda et al. 1997). Hesketh and colleagues have shown that the nuclear translocation is determined by the 3’ untranslated region of MT1 (Hesketh 2004). Recently, they identified a 11 nucleotide sequence in the 3’UTR, containing a CACC repeat, that is necessary for the nuclear translocation of MT1 (Nury et al. 2005).

In the human and mouse MT1 promoters, potential binding sites for the E2F transcription factor are present. This transcription factor is a critical determinant of the G1/S-phase transition during the mammalian cell cycle, serving to activate the transcription of a group of genes that encode proteins necessary for DNA replication. In addition, E2F activity appears to be directly regulated by the action of retinoblastoma protein (Rb). Human DP-1 and E2F-1 associate both in vivo and in vitro, and this interaction leads to enhanced binding to E2F DNA-binding sites (Helin et al. 1993). The association of E2F-1 and DP-1 leads to co-operative activation of an E2F-responsive promoter. It was also demonstrated that trans-activation by E2F-1/DP-1 heterodimers is inhibited by RB. Nevertheless, the actual binding of E2F to MT promoters needs to be determined.

The nuclear need for MT at specific stages of the cell cycle might point towards a critical function of MT in regulating metalloproteins, or protection from DNA damage and apoptosis (Meplan et al. 1999).

**Development.** Metallothionein expression is tightly regulated and activated during mammalian embryonic development. During early development of the mouse embryo, expression of MT1 is induced specifically in the endoderm cells of the visceral yolk sac (Andrews et al. 2001). It was shown that MTF1 is absolutely essential for up-regulation of MT1 gene expression in visceral endoderm cells and that optimal expression also involves the binding of USF to the promoter. Only the human MT1G promoter contains an E box, therefore, it is not known whether the human USF binding is necessary in development.

**Inhibition of transcription**

Contrary to the activation of MT expression, their down-regulation by cis-acting events has not been extensively studied. However, a reduced expression has been frequently found in many types of cancers and in IBD. Three factors are probably involved in suppression of MT: nuclear factor 1 (NFI), ZBTB11 and EGFR-specific transcription factor (ETF). Overexpression of NFI in human hepatoma cells suppressed both constitutive and metal induced activation of the MT1 promoter (Majumder et al. 2001). NFI binds to an MRE-c’ sequence in the mouse promoter
(Datta and Jacob 1993). NFI sites are present in the mouse and all human promoters except MT1B. One study reports a 120 kDa zinc finger protein (PZ120) repressing the transcription of the human MT2A gene by binding to its transcription initiation site (Tang et al. 1999). The PZ120 gene is now replaced in the NCBI database as the zinc finger and BTB domain containing 11 (ZBTB11). So far, no additional reports describing this transcription factor were published since. ETF is present in all MT isoforms, and has not been described in the context of MT inhibition. Overexpression of this factor in primate kidney CV1 cells showed that it represses expression originating from both the EGFR and beta-actin gene promoters (Kageyama and Pastan 1989).

The influence of environmental factors on MT expression was illustrated by the inhibition of zinc induced MT induction by chromium, a major environmental carcinogen (Majumder et al. 2003). This inhibition was working through interfering with MTF1. Treatment of mammalian cells with cycloheximide, a protein synthesis inhibitor, resulted in increased MT1 transcription (McCormick et al. 1991). Recently, the presence of a labile inhibitor of MT1 expression was suggested (Bi et al. 2006). This repressor negatively controls agonist-induced turnover of the MTF1 protein.

**Promoter methylation**

Methylation of cytosine residues in promoter sequences is generally associated with a low transcriptional level of the respective gene. In mammals, methylation is mainly found within so-called CpG islands, regions of high CG content, thought to be involved in transcriptional regulation. The general rule is that CpG islands are not methylated, except for genes on the inactive X chromosome and at imprinted loci. Generally, to study the methylation status of a gene in a cell line or in tissue, its expression is correlated to the amount of methylated CG dinucleotides in genomic DNA isolated from the same source. If a correlation is found, e.g. low gene expression and a hypermethylation of the CpG island, the cells or tissue specimens are subjected to a demethylation agent, such as 5-azacytidine. If the demethylation results in an increased expression of the gene, a role for methylation in expression regulation of the gene is established.

Rodent and human MT genes contain a CpG island in their promoter. Tissue specific methylation of MT1B has been demonstrated (Heguy et al. 1986). The MT1B gene is only expressed in human hepatoma and renal carcinoma cell lines, and not in HeLa cells, where the 5' flanking region of MT1B is highly methylated. MT3 hypermethylation has been shown in gastric cancer (Deng et al. 2003) and in oesophageal squamous carcinoma (Smith et al. 2005).
CONCLUSION

The persistent differential expression of MTs in stress reaction and in pathological conditions suggests a strict regulation of these proteins. It is plausible that a change in e.g. cell cycle regulation of MT could influence cancer development. In addition, cell-type specific expression of MTs may indicate a divergence in functions within cell types or organs. For instance, disability of the tightly controlled proliferation of stem cells in the intestinal crypts may render individuals more susceptible for developing colon cancers. Similarly, a disturbed immune balance in the intestine due to changes in MT expression could be a prerequisite for IBD. Our in silico search for transcriptional regulatory regions in the MT promoters revealed new potential targets that could help to unravel some of the MT regulatory mechanisms. Future studies on the expression and regulation of MT genes are likely to provide insights to their role in both health and disease. Ultimately, novel strategies for manipulating intracellular MT levels could lead to new therapies.

REFERENCES


The regulation of human metallothionein


Chapter 4


-Section II-
SPONDYLOARTHROPATHY AS A MODEL FOR EARLY CROHN’S DISEASE

Debby Laukens

Department of Molecular Biomedical Research, Ghent University and Flanders Interuniversity Institute for Biotechnology (VIB), Gent, Belgium

INTRODUCTION

The spondyloarthropathies (SpA) are a heterogeneous group of chronic inflammatory arthritides that share certain clinical features, and are linked by their association to the human leukocyte antigen class-I gene, \textit{HLA-B27} (Wright 1978). The most characteristic clinical feature is inflammatory back pain, caused by inflammation of the sacroiliac joints (sacroiliitis, Figure 1 left), or the joints of the spinal vertebrae (spondylitis, Figure 1 middle). A specific characteristic of SpA is enthesitis, involving inflammation at the sites where tendons, ligaments or joint capsules are attached to the bone (Figure 1, right). Peripheral joint inflammation is frequently present in SpA, and is mostly pauciarticular (less than four joints), and affects predominantly the lower limbs. These are progressive diseases in which chronic inflammation leads to deterioration of the bone or cartilage. New bone formation in the spine or peripheral joints severely impairs movement in these patients. Usually, low back pain is the first symptom, which is worse at night, in the morning and after periods of inactivity. The course of SpA is highly variable and characterized by spontaneous remissions and flare-ups, particularly in the early stages. Often, extra-articular manifestations are associated with SpA, such as uveitis, psoriasis and inflammatory bowel disease (IBD), either Crohn’s disease (CD) or ulcerative colitis (UC). Since 1991, the European Spondyloarthropathy Study Group (ESSG) criteria are widely used for diagnosis of SpA (Dougados \textit{et al.} 1991). The SpA group is generally divided into 5 main disease categories (Table 1): 1) ankylosing spondylitis (AS), the most common form of SpA; 2)
Table 1 Features of spondyloarthropathies

<table>
<thead>
<tr>
<th>features</th>
<th>Ankylosing spondylitis</th>
<th>Reactive arthritis</th>
<th>Psoriatic arthritis</th>
<th>SpA associated with IBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>prevalence</td>
<td>0.1-0.2%</td>
<td>0.1%</td>
<td>0.2-0.4%</td>
<td>rare</td>
</tr>
<tr>
<td>age at onset</td>
<td>late teens to early adulthood</td>
<td>late teens to early adulthood</td>
<td>35 to 45 years</td>
<td>any age</td>
</tr>
<tr>
<td>male to female ratio</td>
<td>3:1</td>
<td>5:1</td>
<td>1:1</td>
<td>1:1</td>
</tr>
<tr>
<td>HLA-B27</td>
<td>90-95%</td>
<td>80%</td>
<td>40%</td>
<td>30%</td>
</tr>
<tr>
<td>articular manifestations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sacroiliitis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>frequency</td>
<td>100%</td>
<td>40-60%</td>
<td>40%</td>
<td>20%</td>
</tr>
<tr>
<td>distribution</td>
<td>symmetric</td>
<td>asymmetric</td>
<td>asymmetric</td>
<td>asymmetric</td>
</tr>
<tr>
<td>peripheral arthritis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>frequency</td>
<td>occasional</td>
<td>common</td>
<td>symmetric</td>
<td>common</td>
</tr>
<tr>
<td>distribution</td>
<td>asymmetric, lower limbs</td>
<td>asymmetric, lower limbs</td>
<td>very common</td>
<td>asymmetric, any joint</td>
</tr>
<tr>
<td>enthesitis</td>
<td>common</td>
<td>common</td>
<td>common</td>
<td>common</td>
</tr>
<tr>
<td>dactylitis</td>
<td>uncommon</td>
<td>common</td>
<td>common</td>
<td>occasional</td>
</tr>
<tr>
<td>syndesmophotes</td>
<td>delicate, marginal</td>
<td>bulky, nonmarginal</td>
<td>bulky, nonmarginal</td>
<td>uncommon</td>
</tr>
<tr>
<td>extra-articular manifestations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>skin lesions</td>
<td>none</td>
<td>circinate balanitis, keratoderma, blemorrhagicum</td>
<td>psoriasis</td>
<td>erythema nodosum, pyoderma gangrenosum</td>
</tr>
<tr>
<td>nail changes</td>
<td>none</td>
<td>onycholysis</td>
<td>pitting, onycholysis</td>
<td>clubbing</td>
</tr>
<tr>
<td>ocular conditions</td>
<td>acute anterior uveitis</td>
<td>acute anterior uveitis, conjunctivitis</td>
<td>chronic uveitis</td>
<td>chronic uveitis</td>
</tr>
<tr>
<td>cardiac conditions</td>
<td>aortic regurgitation, conduction defects</td>
<td>aortic regurgitation, conduction defects</td>
<td>aortic regurgitation, conduction defects</td>
<td>aortic regurgitation</td>
</tr>
<tr>
<td>oral conditions</td>
<td>ulcers</td>
<td>ulcers</td>
<td>ulcers</td>
<td>ulcers</td>
</tr>
<tr>
<td>pulmonary features</td>
<td>upper lobe fibrosis</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>gastrointestinal conditions</td>
<td>none</td>
<td>diarrhea</td>
<td>none</td>
<td>inflammatory bowel disease</td>
</tr>
<tr>
<td>genitalary conditions</td>
<td>prostatitis</td>
<td>urethritis, cervicitis</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>renal conditions</td>
<td>amyloidosis, IgA nephropathy</td>
<td>amyloidosis</td>
<td>amyloidosis</td>
<td>nephrolithiasis</td>
</tr>
</tbody>
</table>
Spondyloarthropathy as a model for early Crohn’s disease

Figure 1 Radiographs of (left) the pelvis in a patient with sacroiliitis showing sclerosis and erosion of the sacroiliac joints (arrows); (middle) the lumbar spine in a patient with ankylosing spondylitis with complete ossification of the annulus fibrosus; (right) the heel in a patient with periosteal reaction at the plantar fascia insertion and at the Achilles tendon insertion on the calcaneus.

reactive arthritis (ReA), in patients with a recent history of a urogenital or an intestinal infection; 3) psoriatic arthritis (PsA); 4) SpA associated with IBD (SpA-IBD); and 5) undifferentiated spondyloarthropathy (USpA), grouping patients that fulfil the criteria for SpA but are not classifiable in one of the former groups.

There is an important overlap between the different disease manifestations of SpA in families. This argues for a genetic predisposition for SpA (Breban et al. 2003). As mentioned above, the SpAs are linked by a common genetic risk factor, HLA-B27, which is much more prevalent in SpA patients than in other rheumatic diseases or healthy controls (Table 1). HLA class I molecules are highly polymorphic membrane glycoproteins, and are specialized antigen-presenting molecules that form stable complexes with antigenic peptides, displaying them for recognition by CD8+ T cells. It is anchored to the cell membrane by a short transmembrane segment, and consists of three α domains, non-covalently bounded to β2-microglobulin. There are currently 24 genotypic subtypes of HLA-B27 identified, and HLA-B*2705 bears the strongest association to SpA. HLA-B*2702, *2703, *2704, and *2707 are associated with AS. Worth mentioning is that the prevalence of SpA is correlated to the occurrence of HLA-B27. For example, SpA is frequent among Eskimo’s, where the prevalence of HLA-27 is 25 to 40% (Boyer et al. 1994). Conversely, SpA is rare in the Japanese population, where HLA-B27 prevalence is less than 1% (Hukuda et al. 2001). The mechanism by which HLA- B27 confers susceptibility to SpA is not understood. Several hypothesis have been proposed (Kim et al.
2005), including the arthritogenic peptide theory, molecular mimicry and aberrant processing or folding of the heavy chain of HLA-B27.

Although SpA is regarded as one entity of inflammatory arthritides, they display distinct clinical features. It is likely that interplay among genetic and environmental factors is responsible for the various clinical manifestations. Spondyloarthritis is considered to result from infection or exposure to an unknown antigen in genetically susceptible patients. In the case of ReA, a known infection precedes the arthritis by several weeks. This arthritis is thought to be a post-infective phenomenon, rather than resulting from a direct infection, since up to now, no viable bacteria could be cultured from synovial fluid of inflamed joints. This phenomenon is also referred to as a sterile inflammation of the joint, as compared to infectious arthritis caused by direct invasion of the joint space by various micro-organisms. Nevertheless, bacterial antigens were commonly found in the synovium of SpA patients (see next paragraph). Thus, it can never be ruled out that an actual infection has occurred in the patient before the onset of SpA. It has been suggested that the gut is an important portal for antigenic uptake in SpA, and thus plays a pathogenic role in susceptible hosts. Here, we review this remarkable link between the bowel and the joint in SpA.

**INTESTINAL INFLAMMATION, SPONDYLOARTHROPATHY AND IBD**

An interesting link has been found between SpA, intestinal inflammation and IBD (Mielants et al. 2005). In 10 to 20% of IBD patients, inflammatory peripheral arthropathy has been observed, and 7 to 25% have axial involvement (de Vlam et al. 2000; De Vos 2004). Moreover, radiographic sacroiliitis is present in 20 to 25% of IBD patients. Clinically, these forms of arthritis are almost identical to SpA, although they are not associated with *HLA-B27*, indicating a different genetic predisposition. On the other hand, the gut is an important site of inflammation in patients with SpA. In ileocolonoscopic studies of SpA patients, histological signs of gut inflammation were found in more than half of the patients, mostly not presenting any clinical intestinal manifestations, while they were not seen in any other inflammatory joint disease (De Keyser et al. 1998). Remission of the joint inflammation was always linked with a disappearance of the gut inflammation. Two types of inflammation were distinguished: acute inflammation resembling infectious enterocolitis, and chronic inflammation more suggestive of early CD (Cuvelier et al. 1987). In the acute type of inflammation, the mucosal architecture is well preserved. The ileal villi and crypts are infiltrated by polymorphonuclear cells, while in the lamina propria, there is an increased number of inflammatory cells. The chronic type of inflammation is characterized by a clearly disturbed mucosal architecture, and is mostly
Spondyloarthropathy as a model for early Crohn’s disease

indistinguishable from CD. The villi are blunted and fused. The crypts are distorted and the lamina propria is edematous and infiltrated by mononuclear cells. Basal lymphoid follicles occur. In some cases, aphthoid ulcers and granulomas are present. The clinical significance of CD-like alterations in the bowel of SpA patients was shown in a prospective long-term study in which SpA patients were reviewed after several years (Mielants et al. 1995; De Vos et al. 1996). About 13% of patients who showed subclinical chronic gut inflammation on biopsies at the first investigation, developed full-blown CD. This supports the concept of subclinical CD in a subpopulation of SpA patients. Hence, SpA and CD could be regarded as a result of a common inflammatory pathway leading to phenotypes with comparable clinical and pathogenic features.

A number of studies have been performed to provide molecular and genetic support for the clinical observation of the association between gut and joint inflammation in SpA. These are discussed in the next paragraphs.

Genetic arguments for a joint-gut axis

Several studies indicated the specific genetic contribution of intestinal inflammation in SpA patients. Bjarnason and co-workers assessed the presence and inheritance pattern of subclinical intestinal inflammation in first-degree relatives of patients with AS (Bjarnason et al. 2003). They appeared to have an inherited abnormality that leads to subclinical intestinal inflammation, suggesting that this feature is transmitted and thus genetically determined. Detailed genetic analysis to identify risk factors for intestinal inflammation still needs to be done.

We recently performed a prospective clinical and radiological evaluation of 102 CD patients, and found an association between CD-associated CARD15 polymorphisms (Chapter 1) (Hugot et al. 2001; Ogura et al. 2001) and the presence of sacroiliitis: 78% of patients with sacroiliitis were carriers of a mutation in CARD15, compared to 48% of patients without sacroiliitis (Peeters et al. 2004). Here, the carriage of CARD15 mutations is predisposing for the onset of both chronic gut inflammation and sacroiliitis. The molecular mechanism of this association has yet to be investigated. However, replication of these results failed in a multicentre study, so caution must be taken in interpreting the involvement of CARD15 in sacroiliitis.

The most convincing evidence for a joint-gut axis in SpA comes from the HLA-B27 transgenic rats. These animals develop an illness similar to SpA, with manifestations including sacroiliitis, enthesitis, arthritis, skin and nail lesions, ocular inflammation, cardiac inflammation, and inflammation of the gastrointestinal and male genitourinary tracts (Taurog et al. 1994). Furthermore, the severity of the clinical disease correlates with the number of copies of HLA-B27 expressed in the transgenic animal. This model provides direct evidence that over-
expression of the major risk factor in human SpA is causative for both arthritis and bowel inflammation, and supports the existence of an interplay between the gut and the joints. Moreover, if these HLA-B27 transgenic rats are raised in a germ-free environment, they do not develop clinical disease. Once introduced to a regular environment and exposed to bacteria, the rats develop clinical manifestations of SpA, suggesting that commensal flora and/or pathogens are required in the disease onset. In humans, HLA-B27 is strongly associated with AS and ReA (Table 1), and it is plausible that this protein can be pathogenically related to the rat model. Nevertheless, SpA with associated IBD is not highly correlated to HLA-B27: only 30% of this group of patients is positive for HLA-B27. In fact, of the HLA-B27 negative SpA patients, ReA or USpA was diagnosed (Mielants et al. 1993). One explanation might be that there are two forms of SpA related with IBD, one in which HLA-B27 is causative for both joint and gut inflammation, and one where another risk factor is involved. Genetically, it is possible that polymorphisms in another gene, located near HLA-B27, are associated with SpA, as it is believed that transmission of genetic information is confined to haplotype blocks, segments of DNA in which the chance of recombination is very low (Daly et al. 2001). On the other hand, the presence of HLA-B27 proteins might not be sufficient to disturb both joint and gut inflammatory pathways, a feature that is maybe not involved in rodents. Moreover, because the frequency of HLA-B27 in the population-at-large far exceeds that of SpA, suggests that there are more genetic determinants for SpA (Reveille 2004). The search for new genes that are both involved in IBD and SpA is a challenging prospect for the future.

**Immunologic arguments for a joint-gut axis**

The immune infiltrate in the gut mucosa of patients with SpA is significantly different from that of healthy controls, even in the absence of inflammation, suggesting the presence of early inflammatory changes in the intestine of SpA patients. Lymphoid follicles in the intestine are specialized microenvironments where naive T cells respond to specific antigens encountered in the intestinal lumen. Under normal conditions, the number of follicles is limited. However, in the presence of inflammatory stimuli, their number increases. Similar to CD, the number of follicles is increased in the ileum and colon of SpA patients with microscopical normal gut biopsies (Demetter et al. 2002). The amount of leukocytes expressing CD11c was augmented in the ileum of SpA patients, while in the colon CD11a and VCAM-1 expressing leukocytes and CD68^+^ macrophages were higher as compared to controls. The increase in CD68^+^ macrophages and adhesion molecules was also seen in CD (Bernstein et al. 1998). Moreover, the expression of αEβ7 integrin, a specific gut homing receptor for effector T cells, was up-regulated on gut
mucosal T cell lines isolated from SpA patients in the absence of inflammation (Elewaut et al. 1999; Van Damme et al. 2001c). In addition, an up-regulation of E-cadherin, the ligand for αEβ7, and its associated catenins, crucial for intercellular adhesion in epithelial cells, has been demonstrated in clinically overt IBD (Demetter et al. 2000). These observations are indicative of an increased infiltration of T cells and macrophages in the epithelial lining of the gut of SpA patients. Macrophages carrying the scavenger receptor CD163 were increased in colonic mucosa of CD and SpA patients, but not in UC (Demetter et al. 2005). They were also increased in the synovium of SpA patients, where they correlated with inflammatory parameters (Baeten et al. 2002). These CD163+ macrophages produce interleukin 1 (IL1) and tumor necrosis factor alpha (TNF), but not IL10 following lipopolysaccharide (LPS) challenge (Baeten et al. 2002). Thus, an imbalance in cytokine production might lead to chronic inflammatory process in the gut mucosa of SpA and CD patients. Finally, there is a similar predominance of T helper type 1 (Th1) producing mucosal T cells in both SpA and CD, most surprisingly with a proportional decrease of interferon gamma and IL2 producing lymphocytes (Van Damme et al. 2001a; Van Damme et al. 2001b).

Overall, these observations point to the presence of early immune changes in the gut of patients with SpA, and an increased antigen handling and presentation, comparable to changes seen in CD.

**Therapeutical intervention is effective for treatment of both joint and gut inflammation**

Therapeutic intervention in SpA and CD can have a positive influence on both the articular and the intestinal inflammation. Sulfasalazine, which has been successfully used to treat colonic inflammation in UC and CD, was effective in the treatment of the peripheral arthritis accompanying SpA, especially if intestinal inflammation is present (Mielants et al. 1996).

Therapy based on a chimeric monoclonal antibody to TNF, infliximab, was highly effective for the treatment of CD. A pilot study conducted in four patients with CD as well as SpA showed a remission of gut inflammation and a significant improvement of articular and axial symptoms (Van den Bosch et al. 2000). Based on these initial findings, the use of infliximab was explored in a number of studies in patients with different forms of active SpA, and high success rates were reported. Based on these findings, two double-blind, placebo controlled trials were conducted (Braun et al. 2002; Van Den Bosch et al. 2002). A fast and significant improvement of disease was shown in these patients.

Although therapeutical improvement of both joint and gut inflammation is not a valid proof of a direct link between the two sites, it is however a nice example of how clinical observations led to
a significant improvement of treatment in these patients.

**MECHANISM OF JOINT-GUT AXIS**

The exact mechanism that links gut inflammation and joint inflammation is not completely understood. The main hypothesis states that exogenous factors, probably bacterial antigens, are permitted to enter the body through the gut. Bacterial antigens and LPS have been found in the synovial fluid of patients with ReA (Granfors et al. 1989; Granfors et al. 1990). In addition, evidence for the presence of bacterial DNA was shown by PCR in synovial fluid of patients with SpA (Pacheco-Tena et al. 2001). The delivery of bacterial antigens to the joint may trigger and perpetuate local inflammation, and is consistent with the parallelism of flare-ups of joint and intestinal inflammation. How these factors reach the joints is not known, given the fact that no live bacteria can be found. A role for T cells, HLA-B27 and macrophages has been postulated. Increased priming of T cells in the SpA gut fits with the concept that intestinal T cells are involved in the induction of arthritis in ReA (Gaston 1993). Arthritis might arise from a T cell mediated immune response to bacterial antigens and degradation products circulating from the gut to the joint. A proof of concept for the T cell re-circulation hypothesis was put forward by the identification of similar clonal T cell expansion in the colonic mucosa and synovium of a patient with enterogenic SpA (May et al. 2000). In addition, *Yersinia* specific antigenic proliferation of T cell clones, isolated from the synovial fluid of a *Yersinia* triggered ReA patients has been shown (Hermann et al. 1989). This could mean that the same antigenic compounds are present in the joints and the gut of these patients. On the other hand, it was shown that gut-derived leukocytes from patients with IBD bind well to venules in the synovial membrane (Salmi et al. 2001), suggesting that intestinally activated T cells have the capacity to enter the synovium. So far, it is not clear whether the presence of similar T cell clones in joint and gut is due to the presence of identical antigens or because the same set of homing molecules occur at both sites. One can speculate that if similar antigenic peptides are present in the gut and the joint, a sustained HLA-B27 mediated T cell response can trigger inflammation.

Macrophages play a central role in the innate immune recognition of bacterial products. The finding that a similar subset of CD163+ macrophages was found in both the synovium and gut mucosa of SpA patients could reflect a similar pathophysiological condition leading to the influx or maturation of this particular macrophage subset at the two sites. It is surprising that stimulation of these cells with LPS does not induce IL10 secretion, while stimulation with the natural CD163 ligand, haptoglobin-hemoglobin complexes, does actually lead to an anti-
Spondyloarthropathy as a model for early Crohn’s disease

inflammatory response of these cells (Ugocsai et al. 2006). It seems plausible that these cells may induce a similar dysregulation of the cytokine imbalance in the colon and synovium.

CONCLUSION

The ample clinical, immunological and genetic data support the existence of a joint-gut connection, explaining the important role of the gut in the pathogenesis of SpA. Maybe, these observations must not be separated from the involvement of other mucosal sites within the concept of SpA, e.g. uveitis and psoriasis. A similar mechanism might be involved in these clinical manifestations. Ultimately, knowledge of molecular events in these processes is important for classification, diagnosis and therapy of SpA patients.

REFERENCES


Chapter 5

2860-5.


Elewaut, D., N. Van Damme, D. Baeten, et al. (1999). "Intestinal mucosa of patients with spondyloarthritis is enriched with T cells carrying αEβ7 integrin, even in the absence of histologically defined inflammation." Gastroenterology 116: G3069.


Chapter 6

CARD15 GENE POLYMORPHISMS IN PATIENTS WITH SPONDYLOARTHROPATHIES IDENTIFY A SPECIFIC PHENOTYPE PREVIOUSLY RELATED TO CROHN’S DISEASE


Debby Laukens¹, Harald Peeters², Denis Marichal², Bert Vander Cruyssen³, Herman Mielants³, Dirk Elewaut³, Pieter Demetter⁴, Claude Cuvelier⁴, Marthe Van Den Berghe³, Pieter Rottiers¹, Eric M Veys³, Erik Remaut¹, Lothar Steidler¹,⁵, Filip De Keyser³ and Martine De Vos²

¹Department of Molecular Biomedical Research, Ghent University and Flanders Interuniversity Institute for Biotechnology (VIB), Gent, Belgium
²Department of Gastroenterology, Ghent University Hospital, Gent, Belgium
³Department of Rheumatology, Ghent University Hospital, Gent, Belgium
⁴Department of Pathology, Ghent University Hospital, Gent, Belgium
⁵Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland

**ABSTRACT**

*Background & aims:* Association between spondyloarthopathy (SpA) and Crohn’s disease (CD) is a well-known phenomenon. A risk for evolution to CD was already demonstrated in the subgroup of SpA patients with associated chronic gut inflammation. We investigated whether the reported polymorphisms in the *CARD15* gene, a susceptibility gene for CD, are associated with the presence of preclinical intestinal inflammation observed in SpA.

*Methods:* We included 104 SpA patients who underwent an ileocolonoscopy with biopsies between 1983 and 2004. Using RFLP-PCR, we assessed the prevalence of three single nucleotide polymorphisms in the *CARD15* gene (R702W, G908R and 1007fs) and compared them to an ethnically matched CD population and a control population.

*Results:* The carrier frequency of R702W, G908R or 1007fs variants in the SpA populations (20%) was similar as in the control population (17%), but increased to 38% in the subgroup of
Chapter 6

SpA patients with chronic gut inflammation. This was significantly higher than in the other SpA subgroups \( (P=0.001) \) and the control group \( (P=0.006) \) but not significantly different from the prevalence in CD \( (49\%) \). This indicates that \textit{CARD15} polymorphisms are associated with a higher risk for development of chronic gut inflammation.

\textbf{Conclusions:} \textit{CARD15} gene polymorphisms clearly identify a subgroup of patients with SpA associated with chronic intestinal inflammation.

\textbf{INTRODUCTION}

The spondyloarthropathies are a group of interrelated inflammatory diseases characterized by a pauciarticular, peripheral, asymmetrical arthritis and/or axial involvement with ankylosing spondylitis (AS) as prototype (Lawrence \textit{et al.} 1998; Khan 2002). Reported prevalences of spondyloarthropathies vary between 0.2 and 1.9\% (Braun \textit{et al.} 1998). Although association with \textit{HLA-B27} is strong, recent genetic studies suggest a polygenic model of susceptibility (Brown \textit{et al.} 1997; Said-Nahal \textit{et al.} 2000; Reveille \textit{et al.} 2001; Granfors \textit{et al.} 2002).

In up to 60\% of spondyloarthropathy (SpA) patients, articular involvement is associated with subclinical histological evidence of chronic or acute gut inflammation in ileum or colon (De Vos 1989; Simenon \textit{et al.} 1990; Leirisalo-Repo \textit{et al.} 1994). We described a long-term evolution to overt Crohn’s disease (CD) in 13\% of patients with initial chronic gut inflammation (Mielants \textit{et al.} 1995a; De Vos \textit{et al.} 1996). The presence of chronic intestinal inflammation did not relate to \textit{HLA-B27}, but a weak association was found with \textit{HLA-B62} (Mielants \textit{et al.} 1995a).

The observed immunologic similarities in SpA with gut inflammation and CD support the concept that this subgroup of SpA patients can be considered as a model for early immune alterations related to CD. An enrichment of gut mucosal T cell lines with \( \alpha \mathrm{E} \beta 7 \) integrin and an increased expression of its ligand, E-cadherin, was found in intestine of CD as well as SpA patients (Elewaut \textit{et al.} 1998a; Demetter \textit{et al.} 2000; Demetter \textit{et al.} 2002). Re-circulation of T cells primed in the gut to synovial tissue is one potential mechanism by which gut and synovial inflammation could be linked. This hypothesis is supported by an altered expression of \( \beta 7 \) integrins, which are highly expressed within the gut, on synovial T cells from SpA patients compared to rheumatoid arthritis (Elewaut \textit{et al.} 1998b). Another potential mechanism includes trafficking of antigen presenting cells between gut and joint. Consistent with this was the augmented infiltration of gut mucosa and synovium with CD163\(^+\) macrophages (producing IL1 and TNF) in both CD and SpA patients (Baeten \textit{et al.} 2002; Demetter \textit{et al.} 2005). Finally, a comparable beneficial clinical effect of infliximab, a monoclonal antibody to TNF suggests a
CARD15 in spondyloarthropathy

A key role of this cytokine in both diseases (Van den Bosch et al. 2000; Van Den Bosch F 2002). In 2001, a correlation was reported between polymorphisms in the CARD15 gene and an increased susceptibility for CD (Hampe et al. 2001; Hugot et al. 2001; Ogura et al. 2001a). Three independent single nucleotide polymorphisms (SNPs) in CARD15 are associated with CD in about 30 to 46% of patients (1 frameshift mutation, 1007fs (SNP13), and 2 missense mutations, R702W (SNP8) and G908R (SNP12)) (Hugot et al. 2001; Esters et al. 2004). These variants increase the risk for CD by a factor of 3 for heterozygous and by a factor of 38 or 44 for respectively homozygous or compound heterozygous individuals (Hugot et al. 2001). Lower prevalences have been described in CD patients in Scotland, Ireland and Northern Europe, whereas no association could be found in Japan (Inoue et al. 2002; Yamazaki et al. 2002; Helio et al. 2003; Arnott et al. 2004).

CARD15 encodes for an intracellular protein, which is expressed in monocytes, granulocytes, dendritic, epithelial and paneth cells, and has binding affinity for bacterial cell wall components like muramyldipeptides (Girardin et al. 2003). The CARD15 protein is involved in NFκB activation and in apoptosis by two N-terminal Caspase Recruitment Domains (hence the term CARD), although its precise pathogenic role in CD remains to be determined (Ogura et al. 2001b; Bonen et al. 2003; Girardin et al. 2003).

CARD15 gene polymorphisms have also been linked with another related syndrome, Blau’s syndrome, characterized by granulomatous inflammation of uvea, skin and joints (Miceli-Richard et al. 2001).

Several studies have been performed to investigate the role of CARD15 polymorphisms in SpA. These studies did not demonstrate an association with SpA or AS in particular (D’Amato 2002; Miceli-Richard et al. 2002; Breban et al. 2003; Ferreiros-Vidal et al. 2003; van der Paardt et al. 2003). Yet an increased prevalence of CARD15 polymorphisms was found in psoriatic arthritis but not in psoriatic skin disease (Borgiani et al. 2002; Rahman et al. 2003; Young et al. 2003). A recent Italian study however could not confirm this association (Giardina et al. 2004).

Nevertheless, this finding could emphasize the importance of investigating the possible role of these genetic variants in specific, clinical subpopulations of patients. In CD as well, CARD15 polymorphisms seem to be related with certain clinical phenotypes (Abreu et al. 2002; Ahmad et al. 2002; Cuthbert et al. 2002; Lesage et al. 2002; Peeters et al. 2004).

In view of the apparent correlation between gut inflammation in SpA and clinical evolution to CD, we investigated whether the presence of polymorphisms in this susceptibility gene for CD would be associated with gut inflammation in SpA patients.
MATERIALS AND METHODS

Study population
This study included 104 Caucasian SpA patients (according to the ESSG criteria (Dougados et al. 1991), who underwent an ileocolonoscopy with concomitant ileal and colonic biopsies between 1983 and 2004. This population consisted of 74 male and 30 female patients with a mean age of 46 years (range: 21-77 yrs). SpA patients were systematically referred by the rheumatologist for an ileocolonoscopy with biopsies, independent of the presence of GI symptoms. Patients with the diagnosis of clinical Crohn’s disease or psoriasis prior to the diagnosis of SpA, were excluded from the study.

A subgroup of 54 patients, all having a long-term follow-up since their diagnosis of SpA (ranging from 17 to 49 years), was recently clinically reassessed. New follow-up colonoscopies were not performed.

The total SpA population consisted of 75 patients with ankylosing spondylitis (AS) according to the modified New York criteria (van der Linden et al. 1984) and 29 patients with an undifferentiated form of SpA (uSpA). Eighteen AS patients only had axial involvement, whereas 57 AS patients also had peripheral disease (defined as the history or presence of peripheral arthritis and/or enthesitis). Twenty-five uSpA patients had peripheral disease and 4 uSpA patients only had axial involvement. These 4 patients had inflammatory low back pain and fulfilled the ESSG criteria, however not the modified New York criteria for AS.

HLA-B27 status was known in a total of 81 patients. In 53 patients both HLA-B27 and HLA-B62 status were known.

A population of 156 consecutive patients with proven CD on clinical, endoscopical and histological grounds was included as well. This cohort included 57 male and 99 female patients with a mean age of 38 years (range: 18-80 yrs).

Prevalences were also compared to those observed in a control population including 140 individuals.

The study was approved by the local ethics committee. All patients signed an informed consent.

Histological classification
A classification of histologic lesions was used as reported in previous studies (De Vos 1989; Mielants et al. 1995a; Mielants et al. 1995b; Mielants et al. 1995c; De Vos et al. 1996). Three subgroups were distinguished: patients with normal gut histology, acute and chronic
inflammation (Cuvelier et al. 1987). In acute inflammatory lesions normal architecture was well preserved. A mucosal and epithelial infiltration by neutrophils and eosinophils was found, without a significant increase in lymphocytes. Small superficial ulcers covered with fibrin and neutrophils overlying hyperplastic lymphoid follicles were occasionally observed. The lamina propria was oedematous and hemorrhagic and contained mainly polymorphonuclear cells. The pattern of inflammation was similar to that seen in acute self-limiting bacterial enterocolitis. The principal features of chronic inflammatory lesions were mucosal architectural alterations with crypt distortion and atrophy in the colon and villous blunting and fusion in ileal mucosa. Both in ileum and colon there was an increased mixed cellularity and formation of basal lymphoid aggregates in the lamina propria. Whenever one of several biopsies featured chronic lesions, regardless of acute or active inflammation in other fragments, a diagnosis of chronic inflammation was made. Although NSAID may induce intestinal disorders, we and others excluded these drugs as aetiology of reported chronic inflammation (De Vos 1989; Simenon et al. 1990; Altomonte et al. 1994).

**CARD15 genotyping (R702W, G908R and 1007fs), HLA-B27 and HLA-B62 typing**

Genomic DNA was extracted from whole blood using Qiagen blood and cell culture DNA kit (Westburg BV, Leusden, The Netherlands) and genotyped all patients for R702W, G908R and 1007fs using RFLP-PCR, followed by separation of the DNA fragments on a 2.5% agarose gel. The missense mutation R702W (GenBank accession number G67950) abolishes the restriction site for MspI, resulting in an intact 130-bp band for mutant alleles compared to two bands of 54- and 76-bp for wild type alleles (forward primer: 5'-CAG CCC TGA TGA CAT TTC TCT T-3', reverse primer: 5'-AGC CGC TCC TCC TGC ATC TCG TA-3'). The missense mutation G908R (GenBank accession number G67951) creates a restriction site for HinP11. The frameshift mutation 1007fs (GenBank accession number G67955) creates a restriction site for NlaIV. The presence of a mutant allele results in two bands of 219 and 41 bp, while the wild type allele produces a single 260-bp product (forward primer: 5'-CTG AGC CTT TGT TGA GC-3', reverse primer: 5'-TCT TCC AAC CAC ATC CCC ATT-3'). In the patients with a known HLA-B27 and HLA-B62 status, typing of these markers had been performed using the micro-lymphocytotoxicity test according to Terasaki and McClelland (Cuthbert et al. 2002).
Chapter 6

Statistical analysis
Statistical significance was determined by the $\chi^2$ test and Odds Ratio using SPSS (SPSS inc., Chicago, Illinois). Multivariate analysis (logistic regression) was performed to investigate whether an association, found through univariate analysis, was independent from other genetic markers. $P$ values less than 0.05 were considered significant.

RESULTS
We subdivided our cohort in three groups according to the gut histology. Forty patients (38%) had a normal histology, 24 patients (23%) had acute gut inflammation and 40 (38%) showed chronic gut inflammation (Table 1).

Table 1 Prevalence of CARD15 variants in the populations, according to subtypes defined at baseline

<table>
<thead>
<tr>
<th>Classification</th>
<th>N</th>
<th>Carriers of CARD15 variant(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>control</td>
<td>140</td>
<td>24 (17%)</td>
</tr>
<tr>
<td>crohn</td>
<td>156</td>
<td>77 (49%)*</td>
</tr>
<tr>
<td>spondyloarthropathy (SpA)</td>
<td>104</td>
<td>21 (20%)§</td>
</tr>
<tr>
<td>ankylosing Spondylitis (AS)</td>
<td>75</td>
<td>16 (21%)‡</td>
</tr>
<tr>
<td>undifferentiated SpA (uSpA)</td>
<td>29</td>
<td>5 (17%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gut histology in SpA population</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>chronic inflammation</td>
<td>40 (38%)</td>
<td>15 (38%)‡</td>
</tr>
<tr>
<td>acute inflammation</td>
<td>24 (23%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>normal histology</td>
<td>40 (38%)</td>
<td>6 (15%)</td>
</tr>
</tbody>
</table>

*Chi-square: $P<0.001$ (carrier frequency in CD vs control population)
§Chi-square: $P=0.5$ (carrier frequency in general SpA vs control population)
‡Chi-square: $P=0.001$ (chronic inflammation in patients with CARD15 variant vs. chronic inflammation in those without CARD15 polymorphisms)

Univariate analysis
Prevalence of CARD15 polymorphisms in the SpA, CD and control populations. The prevalences of CARD15 polymorphisms in the total SpA (20%), specific AS (21%) and uSpA (17%) population did not differ significantly (Table 1). All except one (homozygous for the 1007fs allelic variant) were heterozygous for at least one mutation. The prevalence of R702W, G908R and 1007fs allelic variants in this SpA population was 12%, 4% and 5% respectively (Table 2). No compound heterozygosity was found. All carriers of CARD15 polymorphisms in the SpA group had (a history of) peripheral disease (Table 3). There were no significant differences
Concerning the disease duration and the duration of the follow-up period between the SpA patients carrying CARD15 polymorphisms and the group of patients without these polymorphisms (data not shown).

In the CD population, a carrier frequency of 49% (77 of 156 patients) was observed (Table 1). Forty-three CD patients carried at least one R702W polymorphism, 14 patients carried at least one G908R polymorphism and 27 patients carried at least one 1007fs polymorphism. Fourteen patients carried two polymorphisms of which 7 patients were homozygous and 7 patients compound heterozygous (Table 2).

In the control group, 24 individuals (17%) carried CARD15 polymorphisms (Table 1). All except one (compound heterozygous for the R702W and 1007fs variant) were single heterozygous (Table 2).

The prevalence of polymorphisms in the SpA cohort (20%) was not different from that observed in the control group (17%) (OR: 1.22, CI: 0.64–2.34, P=0.5) and significantly lower compared to the prevalence found in our CD population (49%) (OR: 3.85, CI: 2.17-6.83, P<0.001).

Table 2 Carrier frequency of CARD15 variants in patients with SpA, CD and controls (%)

<table>
<thead>
<tr>
<th></th>
<th>SpA (N=104)</th>
<th>CD (N=156)</th>
<th>Controls (N=140)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R702W</td>
<td>G908R</td>
<td>1007fs</td>
</tr>
<tr>
<td>CARD15+/-</td>
<td>12 (12)</td>
<td>4 (4)</td>
<td>4 (4)</td>
</tr>
<tr>
<td>CARD15-/-</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>

Number of patients carrying R702W, G908R or 1007fs variants
CARD15+/-: heterozygous; CARD15-/-: homozygous
Overall = total number of patients in the group carrying at least 1 variant
* The sum of all allelic CARD15 variants is greater than the overall number of patients at least carrying one variant, since some patients carry 2 different SNPs, thus displaying a compound heterozygous status

Association between CARD15 polymorphisms and intestinal inflammation in SpA patients. The carrier frequency in the subpopulation of SpA patients with chronic gut inflammation was 38% (15 of 40 patients) which was significantly higher compared to the control population (OR: 2.9, CI: 1.33-6.30, P=0.006) and the other SpA populations (OR: 5.80, CI: 2.02-16.68, P=0.001) and not statistically different from that observed in our CD population (49%, OR: 1.62, CI: 0.80-3.31, P=0.2) (Table 1).

Of all SpA patients carrying CARD15 polymorphisms, 71% (15 out of 21 patients) had chronic gut inflammation, 0% acute inflammation and 29% presented with normal histology (Table 1). The only SpA patient carrying two CARD15 variants also had chronic gut inflammation. In
contrast, only 25 out of 83 patients with a wild type genotype (30%) had chronic gut inflammation, 29% acute inflammation and 41% normal histology. Consequently, the presence of CARD15 polymorphisms was associated with a higher risk for development of chronic gut inflammation.

There are no statistically significant differences between the AS and the uSpA group concerning the prevalence of CARD15 polymorphisms in patients with normal (3/29 in AS vs 3/11 in uSpA, OR: 3.3, CI: 0.5-19.4, P=0.3), acute (0/13 in AS vs 0/11 in uSpA) or chronic (13/33 in AS vs 2/7 in uSpA, OR: 1.6, CI: 0.3-9.7, P=0.7) gut inflammation.

In the subgroup of 54 patients who were clinically reassessed, 4 patients evolved from histological chronic gut inflammation towards clinically overt Crohn’s disease. Two of these 4 patients carried CARD15 polymorphisms. The other 22 patients with chronic gut inflammation in this group did not develop clinical CD.

**Association between CARD15 polymorphisms and HLA-B27 in SpA patients.** There was no significant association between the presence of these 2 genetic markers. Six of 34 HLA-B27 negative patients carried CARD15 polymorphisms versus 13 of 47 HLA-B27 positive patients (OR: 1.8, CI: 0.6-5.3, P=0.3).

**Multivariate analysis**

In the subgroup of 53 SpA patients of whom both HLA-B27 and HLA-B62 status were known, logistic regression was performed (with the presence of chronic gut inflammation as dependent variable). This showed that the association between chronic gut inflammation and CARD15 polymorphisms (OR: 17.3, CI: 2.0-152.3, P=0.01) is independent of HLA-B27 (OR: 1.7, CI: 0.5-6.0, P=0.42) and HLA-B62 (OR: 2.5, CI: 0.5-13.0, P=0.28).

**CONCLUSION**

This study describes a novel and remarkably strong association between variants in a host defence gene located on chromosome 16, CARD15, and a chronic form of gut inflammation in patients with SpA. Interestingly, the prevalence of CARD15 polymorphisms in this subgroup of SpA patients was not significantly different from that observed in patients suffering from CD.

Three single nucleotide polymorphisms have been associated with CD (Hampe et al. 2001; Hugot et al. 2001; Ogura et al. 2001a). One variant (1007fs) encodes a truncated protein which results in altered activation of NFκB in response to bacterial stimuli (Ogura et al. 2001b; Bonen
et al. 2003; Girardin et al. 2003). The two other single nucleotide polymorphisms (R702W and G908R) result in an amino acid substitution.

Table 3 Prevalences of CARD15 polymorphisms according to the presence of mainly axial or peripheral involvement in the total SpA group, AS group and uSpA group

<table>
<thead>
<tr>
<th></th>
<th>total SpA group (N=104)</th>
<th>AS group (N=75)</th>
<th>uSpA group (N=29)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CARD15 P=0.006</td>
<td>CARD15 P=0.006</td>
<td>CARD15 P=1.0</td>
</tr>
<tr>
<td></td>
<td>wild type variant total</td>
<td>wild type variant total</td>
<td>wild type variant total</td>
</tr>
<tr>
<td>axial</td>
<td>22  0  22</td>
<td>18  0  18</td>
<td>4  0  4</td>
</tr>
<tr>
<td>peripheral</td>
<td>61  21  82</td>
<td>41  16  57</td>
<td>20  5  25</td>
</tr>
<tr>
<td>total</td>
<td>83  21  104</td>
<td>59  16  75</td>
<td>24  5  29</td>
</tr>
</tbody>
</table>

More recently, several groups assessed the linkage of CARD15 variants in CD to particular clinical phenotypes but the results of these retrospective studies are disparate. The presence of two mutations has been linked to younger age at onset and preferential involvement of small bowel (Lesage et al. 2002). Preference for ileal involvement was also reported by Cuthbert (Cuthbert et al. 2002) and by Ahmad (Ahmad et al. 2002). Prevalence for fibrostenosing disease was dominant in a study of Abreu (Abreu et al. 2002). In these studies, no association of CARD15 variants with extra-intestinal involvement could be retained.

The present study demonstrates a new association between these three CD-associated variants in the leucine-rich regulatory domain of the CARD15 gene and a distinct subgroup of patients with spondyloarthropathies. Similarly to previous reports, the overall prevalence of mutations in SpA patients was not statistically different from the prevalence in our control population (D’Amato 2002; Miceli-Richard et al. 2002; Breban et al. 2003; Ferreiros-Vidal et al. 2003; van der Paardt et al. 2003). However, unlike the previous studies, we identified a distinct clinical subgroup, characterised by the presence of chronic inflammatory gut lesions, with a remarkably high prevalence (38%) of CARD15 polymorphisms, being not significantly different from the prevalence in the CD population (48%) and significantly higher compared to the control population (17%) and the other SpA patients. Previous studies from our group showed that in particular these patients with chronic gut inflammation were at risk for the progression to CD (Mielants et al. 1995a; De Vos et al. 1996).

Striking in the present study, none of the SpA patients with only axial disease carried CARD15 polymorphisms. Carriers of these polymorphisms all had (a history of) peripheral disease. This is in concordance with previous studies, where more chronic gut inflammation could be found in AS patients with peripheral disease compared to strict axial AS patients (Mielants et al. 1995a).
One previous study investigated *CARD15* polymorphisms in AS patients with CD and ulcerative colitis (UC) (Crane *et al.* 2002). It did not show a higher prevalence of *CARD15* variants in AS patients with CD compared to idiopathic AS, AS with UC or healthy controls. However, the low prevalence of *CARD15* variants in the CD population with AS was not compared to the prevalence in a general CD population and it unexpectedly revealed a possible association between the G908R *CARD15* variant and AS patients with ulcerative colitis. Moreover, in a recent study we found an association between *CARD15* polymorphisms and the presence of radiological sacroiliitis in CD patients, unrelated to the *HLA-B27* status of these subjects. These data already pointed at a role for the *CARD15* gene in the link between gut and joint inflammation (Peeters *et al.* 2004).

Our findings confirm the previous reported clinical, therapeutical and immunological links between SpA and CD and provide also genetic proof for the association between both diseases. Since the chronic gut inflammation in the majority of SpA patients remains asymptomatic, this might suggest that *CARD15* polymorphisms could be linked with the development of (subclinical) chronic gut inflammation rather than with CD as such.

The underlying pathogenetic mechanisms that could explain the phenotypic expression of *CARD15* mutations in SpA need to be investigated. *CARD15* encodes a cytosolic protein that could play a role in SpA by interference with transport of antigens by macrophages from mucosal surfaces to the joints (Salmi *et al.* 2001). *CARD15* seems to function as an intracellular receptor for bacterial components, where the C-terminal leucine-rich regulatory domain is crucial for responsiveness. Cellular response to bacterial products was altered in HEK293T cells transfected with expression plasmids containing any of the three SNPs (Ogura *et al.* 2001b; Bonen *et al.* 2003; Girardin *et al.* 2003). Moreover, expression of *CARD15* in myeloblastic and epithelial cells is enhanced by pro-inflammatory cytokines and bacterial components via NFκB (Ogura *et al.* 2001b; Gutierrez *et al.* 2002; Bonen *et al.* 2003). This response is likely to mediate cytokine production including TNF suggesting that up-regulation of *CARD15* may be part of a positive regulatory loop and facilitate the response of the host to pathogens. A genetically determined disturbed handling of bacterial products in the intestinal tract, leading to an altered transport of antigens by T cells to synovial tissue, is an interesting hypothesis that should be investigated in spondyloarthropathy. A further identification and characterisation of inflammatory cells involved in gut and joint inflammation may also lead to new therapeutic targets.

In conclusion, a distinct phenotype associated with the three main CD associated *CARD15* variants is reported in patients with SpA. Our data show that the presence of *CARD15* variants in
SpA patients strongly predisposes to chronic intestinal inflammation, defining a population at risk for evolution to CD. However, the persistence of the subclinical character of the inflammation in a large part of patients may reflect that CD is a multigenic disease or alternatively that the heterozygous carriage of CARD15 polymorphisms predisposes only to a subclinical inflammation.

REFERENCES


Chapter 7

ALTERED GUT TRANSCRIPTOME IN SPONDYLOARTHROPATHY

Ann Rheum Dis. 2006 Feb

Debby Laukens¹, Harald Peeters², Bert Vander Cruyssen³, Tom Boonefaes¹, Dirk Elewaut³, Filip De Keyser³, Herman Mielants³, Claude Cuvelier⁴, Eric M. Veys³, Kenny Knechť⁵, Paul Van Hummelen⁶, Erik Remaut¹, Lothar Steidler¹, ⁷, Martine De Vos² and Pieter Rottiers¹

¹Department for Molecular Biomedical Research, Ghent University and Flanders Interuniversity Institute for Biotechnology (VIB), Gent, Belgium
²Department of Gastroenterology, Ghent University Hospital, Gent, Belgium
³Department of Rheumatology, Ghent University Hospital, Gent, Belgium
⁴Department of Pathology, Ghent University Hospital, Gent, Belgium
⁵Applied Maths BVBA, Sint-Martens-Latem, Belgium
⁶MicroArray Facility, VIB, Leuven, Belgium
⁷Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland

ABSTRACT

Background & aims: Intestinal inflammation is a common feature of spondyloarthropathy (SpA) and Crohn's disease (CD). Inflammation is manifested clinically in CD, and subclinical in SpA. However, a fraction of SpA patients develop overt CD. The aim of this study was to investigate whether subclinical gut lesions in SpA patients are associated with transcriptome changes comparable to those seen in CD. We examined global gene expression in non-inflamed colon biopsies, and screened for differentially expressed genes.

Methods: Macroarray analysis was used as an initial genome-wide screen for selecting a comprehensive set of genes relevant to CD and SpA. This led to the identification of 2,625 expressed sequence tags (ESTs) that are differentially expressed in the colon of CD and/or SpA patients. These clones, together with appropriate controls (6,779 in total) were used to construct a glass-based microarray, which was then used to analyze colon biopsies from 15 SpA patients,
Chapter 7

11 CD patients, and 10 controls.

Results: Ninety-five genes were identified as differentially expressed in SpA patients with a history of subclinical chronic gut inflammation as well as in CD patients. Principal component analysis of this filtered set of genes successfully distinguished colon biopsies from the three groups studied. Spondyloarthropathy patients with subclinical chronic gut inflammation cluster together, and are more related to CD.

Conclusions: The transcriptome in the intestine of SpA patients differs from that of controls. Moreover, these gene alterations are comparable to those seen in CD, confirming initial clinical observations. Based on these findings, new (genetic) markers for detection of early CD in SpA patients can be considered.

INTRODUCTION

The clinical association between Spondyloarthropathy (SpA) and Crohn’s disease (CD) is illustrated by the concurrence of similar arthropathy and intestinal inflammation in the two diseases, indicating a shared etiology and pathogenesis. Depending on the imaging technique used, up to one third of CD patients have peripheral and/or sacroiliac joint abnormalities similar to those seen in various SpA subgroups (Davis et al. 1978; Scott et al. 1990). In addition, 60% of SpA patients who have no evidence of CD exhibit endoscopic and/or histological signs of subclinical gut inflammation (Mielants et al. 1995a). In general, two types of inflammation are observed: acute inflammation as seen in infectious colitis, and chronic inflammation resembling that in CD (Mielants et al. 1995a). A striking parallel exists between the activity of inflammation at the joints and the intestine. Moreover, long-term evolution to CD was observed in 13% of SpA patients with initial chronic gut inflammation, supporting the concept of preclinical CD in those patients (De Vos et al. 1996). Since these clinical observations, several studies provided additional evidence for a joint-gut axis on the molecular as well as the genetic level. The early immune alterations observed are up-regulation of αEβ7 integrin on T-cell lines from SpA patients (Elewaut et al. 1998b), and an increase in lymphoid follicles and lamina propria mononuclear cells in intestinal biopsies (Demetter et al. 2000; Demetter et al. 2002). Increased expression of αEβ7 and the E-cadherin/catenin complex was found in gut mucosa from CD and SpA patients (Elewaut et al. 1998a; Demetter et al. 2000). A specific subset of CD163+ macrophages is augmented in both groups of patients, supporting the hypothesis of a recirculation of similar clones in the intestinal mucosa and synovium (Baeten et al. 2002).

Both CD and SpA are termed complex genetic traits, because many genes are probably involved.
Altered gut transcriptome in spondyloarthropathy

in pathogenesis, and environmental factors have a substantial influence on the outcome of the disease. Evidence exists for a common genetic risk factor in the development of subclinical intestinal inflammation in first-degree relatives of patients with ankylosing spondylitis, which is the prototype of SpA (Bjarnason et al. 2003). Furthermore, we found that CARD15, which was the first CD susceptibility gene identified, is associated with chronic subclinical inflammation in patients with SpA (Laukens et al. 2004). In this regard, SpA patients can serve as a unique model for early CD.

In order to determine whether the association between the two disorders occurs not only at the clinical but also at the transcriptome level, we compared global gene expression in non-inflamed colon biopsies from SpA and CD patients. We propose that it is possible to identify a set of genes that distinguish CD patients and SpA patients with a history of chronic gut inflammation from SpA patients without chronic gut inflammation and from controls.

METHODS

**Patients, tissue collection and histological classification**

Colon biopsies from CD and SpA patients and healthy controls were obtained during colonoscopy. All biopsies were taken from non-inflamed sigmoid at 30 cm. Biopsy specimens were immediately placed in RNAlater (Ambion, Cambridgeshire, UK) and frozen at -80°C until sample processing. Three biopsies were obtained from each of 34 patients diagnosed with CD according to clinical, endoscopic and histological criteria, and 20 patients diagnosed with SpA according to ESSG criteria (Dougados et al. 1991). Sixteen patients without clinical manifestations of CD or SpA, who were undergoing colonoscopy for colon cancer screening, were included as a control population.

Histological classification of the SpA ileum and colon was performed as in our previous studies (De Vos 1989; Mielants et al. 1995a; Mielants et al. 1995b; Mielants et al. 1995c; De Vos et al. 1996). We distinguished three classes: patients with normal histology, patients with acute inflammatory lesions, and those with chronic inflammatory lesions (Cuvelier et al. 1987). In acute lesions, normal architecture was well preserved. There was infiltration by neutrophils and eosinophils without a considerable increase in lymphocytes. Small superficial ulcers covered with fibrin and neutrophils overlying hyperplastic lymphoid follicles were occasionally observed. The lamina propria was edematous and hemorrhagic, and contained mainly polymorphonuclear cells. The pattern of inflammation was similar to that seen in acute self-limiting bacterial enterocolitis. The principal features of chronic lesions were crypt distortion,
atrophy of the villous surface of the mucosa, villous blunting and fusion, increased mixed cellularity, and basal lymphoid aggregates in the lamina propria. Though several biopsies were obtained from each patient, a diagnosis of chronic inflammation was made even if only one biopsy showed chronic lesions, regardless of acute or active inflammation in the other biopsies. SpA patients who had chronic inflammation in colon and/or ileum in previous examinations were termed SpA patients with chronic gut inflammation.

**RNA extraction**

Total RNA was extracted from biopsies using the Qiagen Rneasy Mini Kit (Westburg BV, Leusden, The Netherlands) with on-column DNAse treatment (Qiagen). Needle homogenization was performed. RNA quality and concentration were checked on the Agilent 2100 Bioanalyzer (Agilent Technologies, Germany).

**Macroarray hybridization and analysis**

Colony filters containing 74,828 expressed sequence tag (EST) clones (Human UniGene collection 2, RZPD, Germany) were used as initial screen. Radioactively labeled probes were produced by incorporation of \( \alpha^{33}\text{P} \text{dCTP during reverse transcription of 50 } \mu\text{g total RNA (MMLV, Promega, Leiden, The Netherlands), using oligoT as primer.}^{33}\text{P-cDNA probes were purified on G-50 spin columns (Amersham Biosciences, Roosendaal, The Netherlands). Hybridization was performed at 10}^{6} \text{cpm/ml at 65°C for 20 hours. Images were acquired after 6, 18 and 24 hours of exposure, using a Phosphorimager system (Amersham Biosciences).} \text{Spot definition and intensity measurement was done using Visualgrid (GPC Biotech AG, Munich, Germany). The raw expression data were processed with an in-house algorithm based on MS Access. Spot intensities were corrected for the local background, followed by a quality control of spots to exclude those influenced by intense signals of adjacent spots. The detection limit for expression values above background was calculated based on the variation of the local background intensity. Constitutive genes (those that show the lowest coefficient of variation over all arrays) were used for normalization. Subsequently, quantitative measures of each clone (gene) were calculated by log2 transformation of the ratio of the mean spot intensity of CD or SpA patients to the mean spot intensity of controls.**

**Microarray hybridization, scanning and analysis**

Construction of a focus microarray chip, probe labeling, hybridization, washing and scanning were carried out at the MicroArray Facility of the Flanders Interuniversity Institute for
Biotechnology (MAF, Leuven, Belgium). Clones selected from the macroarray screen were PCR-amplified from RZPD clones using universal M13 primers. PCR fragments were purified on MultiScreen PCR plates (Millipore, Brussels, Belgium) and resuspended in 50% DMSO at an average concentration of 100 ng/µl. The PCR products were arrayed in duplicate on Type VII silane-coated slides using a Molecular Dynamics Generation III printer (Amersham Biosciences, Buckinghamshire, UK). Total RNA (5 µg) was amplified using a modified protocol of in vitro transcription as described (Puskas et al. 2002). Five µg of the amplified RNA were Cy3- or Cy5-labeled as described at http:\www.microarrays.be\service.htm. Arrays were scanned at 532 and 635 nm using a Generation III scanner (Amersham BioSciences). Images were analyzed with ArrayVision (Imaging Research Inc., Ontario, Canada). Each hybridization was repeated in a dye swap experiment. Spot intensities were measured, corrected for local background, and those that exceeded the background by more than two standard deviations (SD) were included. For each gene, ratios of red (Cy-5) to green (Cy-3) intensities (I) were calculated and normalized via a Lowess Fit of the log2 ratios \[\log_2(I_{Cy-5}/I_{Cy-3})\] over the log2 total intensity \[\log_2(I_{Cy-5} \times I_{Cy-3})\].

For comparing the microarray datasets, a mixture of RNA from 5 CD patients, 5 SpA patients and 5 controls was used as reference RNA. This reference sample provides a positive hybridization signal at each probe element on the microarray, which is essential when calculating and comparing fluorescence ratios. The data were imported into GeneMaths® XT (Applied Maths, St-Martens-Latem, Belgium). Weighted mean ratios and their corresponding error (pixel SD) were calculated from the dye swap. Data were normalized over all arrays, and missing values were imputed using k-nearest neighbor algorithm (20 neighbors). GeneMaths® XT was used to perform all subsequent supervised and unsupervised analyses.

Statistics
All P-values chosen for cut-off are subjective.

RESULTS

Design of the custom microarray
In order to provide a practical and cost effective tool for conducting a large number of hybridizations, a self-designed focus microarray chip was constructed specifically for studying colonic gene expression in SpA and CD. To accomplish this, a genome-wide survey of gene expression in colon biopsies of 4 CD patients, 4 SpA patients and 6 controls was conducted
using high-density nylon arrays containing 74,828 cDNA sequences (Table 1, macroarrays). Spots that showed aberrant morphology, encompassed variation in replicates or were impaired because of over shining (characteristic of radioactive signals) were filtered out and considered as clones lost through experimental error. To select for clones that are differentially expressed in CD or SPA patients, we arbitrarily selected for those that have a log₂ transformed mean ratio of less than -0.6 or more than +0.6 (1.5 fold down- or up-regulated). Genes that might be differentially expressed between groups (control versus CD or control versus SpA) were identified using a simple algorithm based on the t-test \((P<0.05)\) and F-values \((P<0.05)\) as selection criteria, providing that at least three consistent intensity values were present in each group. F-values were chosen for selection because we assumed that differences in variances within groups can be important. A total of 2,652 clones were identified as “potentially differentially expressed”. These genes, together with 4,127 ESTs lost through experimental error and which might include, beside control ESTs, additional differentially expressed genes, were used to produce a glass-based microarray platform. This allowed us to screen more patients in a more accurate and sensitive manner.

**Clustering of unfiltered data**

We hybridized an independent cohort of 15 SpA, 11 CD and 10 control patients to the focus microarray (Table 1, microarrays). Unsupervised clustering (without prior knowledge of groups) using all genes revealed no clustering with respect to disease or phenotype (e.g. type of intestinal inflammation). The inability to find discriminatory genes using unfiltered data is perhaps not too surprising, as we are analyzing the steady state transcriptome in non-inflamed tissue samples of complex inflammatory diseases. Subtle differences in only a few genes are lost in the vast number of random variations. The problem of detecting differentially expressed genes can be overcome by performing supervised clustering. To this end we divided the patients into four main groups: CD, SpA with chronic gut inflammation, SpA without chronic gut inflammation, and controls. Discriminant analysis can reduce \(N\)-dimensional data into a more visual 2-D or 3-D plot, with prior knowledge of groups (Figure 1). With this approach, the above-defined groups became clearly separated, indicating that our full dataset contains genes that can differentiate between these disease states.
### Table 1 Study population

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample</th>
<th>Sex</th>
<th>Age</th>
<th>SpA</th>
<th>Gut Histology</th>
<th>Clinical</th>
<th>CD Location</th>
<th>Medication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macroarrays Control</td>
<td>1</td>
<td>F</td>
<td>59</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>F</td>
<td>45</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>F</td>
<td>55</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>M</td>
<td>30</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>F</td>
<td>58</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>M</td>
<td>40</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>7</td>
<td>M</td>
<td>21</td>
<td>no</td>
<td>chronic</td>
<td>yes</td>
<td>IC</td>
<td>5-ASA</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>F</td>
<td>48</td>
<td>no</td>
<td>chronic</td>
<td>yes</td>
<td>IC</td>
<td>5-ASA</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>F</td>
<td>41</td>
<td>yes</td>
<td>chronic</td>
<td>yes</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>F</td>
<td>23</td>
<td>no</td>
<td>chronic</td>
<td>yes</td>
<td>C</td>
<td>AZA</td>
</tr>
<tr>
<td>SpA</td>
<td>11</td>
<td>M</td>
<td>47</td>
<td>yes</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>F</td>
<td>85</td>
<td>yes</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>F</td>
<td>60</td>
<td>yes</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>F</td>
<td>43</td>
<td>yes</td>
<td>chronic</td>
<td>yes</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Microarrays Control</td>
<td>1</td>
<td>M</td>
<td>54</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>F</td>
<td>64</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>F</td>
<td>72</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>M</td>
<td>51</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>F</td>
<td>21</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>F</td>
<td>68</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>F</td>
<td>32</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>F</td>
<td>73</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>F</td>
<td>66</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>M</td>
<td>76</td>
<td>no</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>11</td>
<td>F</td>
<td>23</td>
<td>no</td>
<td>chronic</td>
<td>yes</td>
<td>IC</td>
<td>5-ASA</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>F</td>
<td>39</td>
<td>yes</td>
<td>chronic</td>
<td>yes</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>M</td>
<td>51</td>
<td>no</td>
<td>chronic</td>
<td>yes</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>M</td>
<td>43</td>
<td>no</td>
<td>chronic</td>
<td>yes</td>
<td>I</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>M</td>
<td>36</td>
<td>no</td>
<td>chronic</td>
<td>yes</td>
<td>IC</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>F</td>
<td>27</td>
<td>no</td>
<td>chronic</td>
<td>yes</td>
<td>IC</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>F</td>
<td>46</td>
<td>yes</td>
<td>chronic</td>
<td>yes</td>
<td>IC</td>
<td>AZA</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>F</td>
<td>23</td>
<td>no</td>
<td>chronic</td>
<td>yes</td>
<td>IC</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>M</td>
<td>19</td>
<td>no</td>
<td>chronic</td>
<td>yes</td>
<td>IC</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>F</td>
<td>26</td>
<td>no</td>
<td>chronic</td>
<td>yes</td>
<td>IC</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>F</td>
<td>36</td>
<td>no</td>
<td>chronic</td>
<td>yes</td>
<td>IC</td>
<td>-</td>
</tr>
<tr>
<td>SpA</td>
<td>22</td>
<td>M</td>
<td>40</td>
<td>AS periph</td>
<td>acute</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>M</td>
<td>29</td>
<td>AS periph</td>
<td>normal</td>
<td>no</td>
<td>NSAID</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>M</td>
<td>42</td>
<td>AS periph</td>
<td>acute</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>M</td>
<td>31</td>
<td>AS periph</td>
<td>acute</td>
<td>no</td>
<td>NSAID</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>M</td>
<td>76</td>
<td>uSPA</td>
<td>normal</td>
<td>no</td>
<td>steroids</td>
<td></td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>F</td>
<td>28</td>
<td>AS ax</td>
<td>normal</td>
<td>no</td>
<td>NSAID</td>
<td></td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>M</td>
<td>58</td>
<td>AS periph</td>
<td>normal</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>M</td>
<td>38</td>
<td>AS ax</td>
<td>normal</td>
<td>no</td>
<td>sulfa</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>F</td>
<td>49</td>
<td>AS periph</td>
<td>chronic</td>
<td>no</td>
<td>sulfa</td>
<td></td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>M</td>
<td>49</td>
<td>AS periph</td>
<td>chronic</td>
<td>no</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>M</td>
<td>45</td>
<td>AS periph</td>
<td>chronic</td>
<td>no</td>
<td>sulfa</td>
<td></td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>M</td>
<td>29</td>
<td>uSPA</td>
<td>acute</td>
<td>no</td>
<td>sulfa+NSAID</td>
<td></td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>M</td>
<td>48</td>
<td>AS periph</td>
<td>normal</td>
<td>no</td>
<td>NSAID</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>F</td>
<td>44</td>
<td>AS periph</td>
<td>chronic</td>
<td>yes</td>
<td>IC</td>
<td>sulfa+AZA</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>M</td>
<td>36</td>
<td>AS ax</td>
<td>normal</td>
<td>no</td>
<td>sulfa+NSAID</td>
<td></td>
</tr>
</tbody>
</table>

UspA = undifferentiated SpA; AS ax = ankylosing spondylitis with only axial involvement; AS periph = ankylosing spondylitis with peripheral involvement; histology of SpA patients is a historical classification; IC = ileocolonic; I = ileal involvement only; C = colonic involvement only; 5-ASA = 5-aminosalicylates; AZA = azathioprine; NSAID = non-steroidal anti-inflammatory drugs; sulfa = sulfasalazine
Identification of genes whose differential regulation is common to both SpA with chronic gut inflammation and CD

By using an independent t-test, we identified 123 genes that are expressed differentially between CD and control ($P<0.01$). With this set we were unable to discriminate SpA patients from controls, although three out of four of the SpA patients with chronic gut inflammation clustered together, indicating the presence of alterations similar to those observed in CD. Thus it was logical to screen for genes modulated commonly in CD and control on one hand, and SpA, SpA with chronic gut inflammation and control on the other hand. In order to include a larger number of genes in this analysis, the statistical significance level was lowered from $P<0.01$ to $P<0.05$. This led to the identification of two sets of genes whose expression pattern discriminates CD from control ($P<0.05$, $N=630$) and SpA from control ($P<0.05$, $N=464$). The latter significance level was determined by ANOVA, in which SpA patients with chronic gut inflammation were defined as a distinct group. The set of 95 genes that are differentially expressed in both CD and SpA could distinguish the three disease groups (Figure 2A, Table 2). In addition, SpA patients with chronic gut inflammation cluster together and are more related to the CD cluster than to the control/SpA cluster, but they remain a separate entity (Figure 2A). Principal component analysis using this set of 95 genes, another way of representing the data, clearly discriminates our patient groups (Figure 2B). Based on the identification of a set of genes involved in CD and also implicated in SpA, we attempted to identify genes that might render these individuals more susceptible to develop CD.
Genes within the CD/SpA chronic cluster

Genes whose expression is aberrant in both CD and SpA with chronic gut inflammation are represented in Table 2 (↑: up-regulated; ↓: down-regulated, P<0.05). Among them, two genes had already been described in the context of CD. Acyl-coenzyme A oxidase 1 (ACOX1), which is the first enzyme of the fatty acid beta-oxidation pathway, donates electrons directly to molecular oxygen, thereby producing hydrogen peroxide. The enzymatic activity of ACOX1 was diminished in both inflamed and non-inflamed areas in CD (Aimone-Gastin et al. 1994). Our observation of ACOX1 transcript down-regulation corroborates this report, and indicates a fault at the level of transcription or mRNA stability.

Figure 2 A. Complete Linkage clustering based on a set of 95 genes whose expression is deviant in CD patients and in SpA patients with chronic gut inflammation as compared to healthy controls. Two main clusters mark a SpA/control cluster and a CD/SpA with chronic inflammation cluster. B. Principal Component Analysis-view using a set of 95 genes whose expression is deviant in CD patients and in SpA patients with chronic gut inflammation. CD (green) SpA without chronic inflammation (blue), SpA with chronic inflammation (yellow) and controls (red).
Table 2 Ninety-five EST’s that cluster CD and SpA patients with chronic gut inflammation

<table>
<thead>
<tr>
<th>Accession</th>
<th>Gene description</th>
<th>Chromosomal location</th>
<th>Expression in CD/SpA chronic</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>H81904</td>
<td>acetyl-Coenzyme A synthetase 2 (ADP forming)</td>
<td>20q11.22</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>A1758270</td>
<td>chromosome 6 open reading frame 32</td>
<td>6p22.3-p21.32</td>
<td>↓</td>
<td>(Barmada et al. 2004)</td>
</tr>
<tr>
<td>A1349525</td>
<td>hypothetical protein FLJ34790</td>
<td>17p13.1</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>A1445844</td>
<td>pyruvate dehydrogenase kinase, 3</td>
<td>Xq22</td>
<td>↓</td>
<td>(van Heel et al. 2004)</td>
</tr>
<tr>
<td>H56656</td>
<td>GDP dissociation inhibitor 2</td>
<td>17p13.1</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>N22829</td>
<td>TatD DNase domain containing</td>
<td>8q24.13</td>
<td>↓</td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>H87107</td>
<td>ring finger protein 5</td>
<td>6p21.3</td>
<td>↓</td>
<td>(Rioux et al. 2000)</td>
</tr>
<tr>
<td>A1680066</td>
<td>adenylate kinase 7</td>
<td>14q32.2</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>R17390</td>
<td>sorting nexin 15</td>
<td>11q12</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>AA812701</td>
<td>hypothetical protein LOC340178</td>
<td>6q27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA825971</td>
<td>hypothetical protein LOC340178</td>
<td>2q13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA291593</td>
<td>solute carrier family 39 (metal ion transporter) member 11</td>
<td>17q24.3-q25.1</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>AA420968</td>
<td>chromosome 6 open reading frame 112</td>
<td>6q21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R08643</td>
<td>interferon regulatory factor 2 binding protein 2</td>
<td>1q42.3</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>T83584</td>
<td>mannosidase, beta A, lysosomal</td>
<td>4q22.2-q25</td>
<td>↓</td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>AA481482</td>
<td>E74-like factor 3 (ets domain transcription factor, epithelial-specific)</td>
<td>1q12.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T78477</td>
<td>zinc finger protein 297</td>
<td>6p21.3</td>
<td>↓</td>
<td>(Rioux et al. 2000)</td>
</tr>
<tr>
<td>H27674</td>
<td>DAB2 interacting protein</td>
<td>9q33.1-3q33.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R13194</td>
<td>BRF1 homolog</td>
<td>14q</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R98578</td>
<td>hypothetical protein FLJ32731</td>
<td>8p11.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H96900</td>
<td>MGC2747</td>
<td>19p13.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N62199</td>
<td>kinesin family member 1B</td>
<td>1p32.6</td>
<td>↓</td>
<td>(Cho et al. 2000)</td>
</tr>
<tr>
<td>R51052</td>
<td>S-phase response (cyclin-related)</td>
<td>1q42.11-42.3</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>AI286348</td>
<td>SH3 domain containing ring finger 2</td>
<td>5q32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R17478</td>
<td>proliferation-inducing protein 13</td>
<td>1q25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1916359</td>
<td>secretory pathway calcium ATPase 2</td>
<td>16q24.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H19034</td>
<td>bladder cancer associated protein</td>
<td>20q11.2-q12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N39050</td>
<td>cDNA FLJ39784 fis, clone SPLEN2002314</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R13544</td>
<td>F-box and leucine-rich repeat protein 3A</td>
<td>13q22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI286163</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R07321</td>
<td>Acyl-Coenzyme A oxidase 1, palmitoyl</td>
<td>17q24-17q25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R12168</td>
<td>staufen, RNA binding protein, homolog 2 (Drosophila)</td>
<td>8q13-q21.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H85472</td>
<td>N-myc downstream regulated gene 1</td>
<td>8q24.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA513663</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R97820</td>
<td>zinc finger CCCH type, antiviral 1</td>
<td>7q34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N62837</td>
<td>leukocyte immunoglobulin-like receptor, subfamily A 4</td>
<td>19q13.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA019615</td>
<td>ubiquitin carboxyl-terminal hydrolase L5</td>
<td>1q32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI809092</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA758064</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI219353</td>
<td>cancer susceptibility candidate 2</td>
<td>10q26.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI698801</td>
<td>hematological and neurological expressed 1</td>
<td>17q25.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA468448</td>
<td>histone 1, H2ac</td>
<td>6p21.3</td>
<td></td>
<td>(Barmada et al. 2004)</td>
</tr>
<tr>
<td>H23734</td>
<td>thymosin, beta 4, X-linked</td>
<td>Xq21.3-q22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H75688</td>
<td>sorting nexin 7</td>
<td>1p21.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N63669</td>
<td>triple functional domain (PTPRF interacting)</td>
<td>5p15.1-p14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI819016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI831827</td>
<td>ADP-ribosylation factor guanine nucleotide-exchange factor 2</td>
<td>20q13.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI923299</td>
<td>hypothetical LOC158730</td>
<td>Xp21.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI286257</td>
<td>zinc finger protein 263</td>
<td>16p13.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI863417</td>
<td>runt-related transcription factor 1</td>
<td>21q22.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2 continued

<table>
<thead>
<tr>
<th>Accession</th>
<th>Gene description</th>
<th>Chromosomal location</th>
<th>Expression in CD/SpA chronic</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI274438</td>
<td>AI289775 weakly similar to 0412263A Ig G2 pFc' PIG Gm</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI345173</td>
<td>AI928384</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA490519</td>
<td>clone IMAGE:5286005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI696951</td>
<td>FLJ17163 sps</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI339536</td>
<td>IMAGE:5286812, mRNA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI245190</td>
<td>AI261428</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI809310</td>
<td>heet domain and RCC1 (CHC1)-like domain 1</td>
<td>15q22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI831536</td>
<td>interleukin 12 receptor, beta 2</td>
<td>1p31.3-p31.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T79944</td>
<td>AA480677 IMAGE:4865533</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA009461</td>
<td>MUF1 protein</td>
<td>1p34.1</td>
<td>↑</td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>R61783</td>
<td>synaptogyrin 2</td>
<td>17q25.3</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>AI282992</td>
<td>glutathione peroxidase 2 (gastrointestinal)</td>
<td>14q24.1</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>W65310</td>
<td>AA588676 PDZ and LIM domain 3</td>
<td>10q22.3-23.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI811147</td>
<td>AI577718</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N58523</td>
<td>W74496</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI352320</td>
<td>AI280708 thyminylate kinase family LPS-inducible member</td>
<td>2p25.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI815599</td>
<td>proteasome (prosome, macropain) 26S subunit, non-ATPase, 11</td>
<td>17q11.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI718161</td>
<td>N98921 tropomyosin 3</td>
<td>1q21.2</td>
<td></td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>AI367201</td>
<td>HSPC135 protein</td>
<td>3q13.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA425545</td>
<td>RP42 homolog</td>
<td>3q26.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H44213</td>
<td>hypothetical gene supported by BC031266</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS8936</td>
<td>nuclear receptor coactivator 7</td>
<td>6q22.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N90208</td>
<td>hypothetical protein FLJ32440</td>
<td>8q24.13</td>
<td></td>
<td>(Vermeire et al. 2004)</td>
</tr>
<tr>
<td>AA427403</td>
<td>interferon induced transmembrane protein 1 (9-27)</td>
<td>11p15.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R25648</td>
<td>phosphatidylinositol 4-kinase type-II beta</td>
<td>4p15.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N93265</td>
<td>superoxide dismutase 1, soluble (amyotrophic lateral sclerosis)</td>
<td>21q22.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H99865</td>
<td>nuclear receptor subfamily 2, group F, member 2</td>
<td>15q26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Genes with aberrant expression in both CD and SpA patients with chronic gut inflammation (↓ down-regulated and ↑ up-regulated as compared to the control and SpA population (P<0.05). Genetic markers are cited for genes that are located within or near (5 cM) one of the CD loci.

Glutathione peroxidase 2 (gastrointestinal glutathione peroxidase, giGSH-Px) is one of the four types of selenium-dependent glutathione peroxidases. Its exclusive expression in the gastrointestinal tract indicates that it functions as a barrier against the absorption of dietary hydroperoxides, and as protection against damage from endogenously formed hydroxyl peroxides. Its activity is increased in ulcerative colitis patients in the active and in the remission stages (Beno et al. 1997). In CD, plasma levels of giGSH-Px are increased (Tuzun et al. 2002). We found that this gene is over-expressed in normal colon tissue in CD and SpA patients with a history of chronic gut inflammation, and so it can act as a marker expressed at non-pathological sites in the intestine in CD and CD susceptible SpA patients.
CONCLUSION

Clinical study of intestinal abnormalities in SpA patients has previously relied on cytokine profiles and immunological alterations. In addition to analyzing individual proteins, genome-wide transcript profiles can be analyzed by microarrays. Global gene expression analysis in non-inflamed colon tissue was used to find genes that are differentially expressed in both CD patients and SpA patients with a history of chronic gut inflammation. Previous studies of gene expression in IBD have focused on biopsies of actively inflamed tissues (Lawrance et al. 2001; Dooley et al. 2004; Stoll et al. 2004). The use of samples from non-inflamed areas from CD patients offers the possibility of identifying early markers for CD, which would permit prediction of the evolution to CD in SpA patients. Moreover, changes in the expression of genes that are regulated during inflammation would be more prominent than the subtle alterations in non-inflammatory genes, although it cannot be ruled out that this procedure will also pick up genes whose differential expression is a consequence and not a cause of the disease. Additionally, looking at basal gene expression may allow us to take into account genetic influences, since gene expression is highly heritable (Morley et al. 2004). Therefore, future studies on markers for CD should concentrate primarily on genes that are located near one of the known loci for CD (Table 2). Genes located within a region linked to CD or IBD in general (if multipoint linkage was performed), or within five centimorgan of the markers that are linked to CD or IBD (in case of two point linkage) should be considered first. Using a model for early CD when identifying CD susceptibility genes can circumvent the heterogeneity of the disease, because probably only a very small number of CD genes will be implicated in SpA.

Array analysis is a rapid procedure for studying the expression of many genes in no more than several samples. Because the number of samples is limited, and the number of genes explored is usually large, false-positive results will obviously arise. Nevertheless, array analysis enables us to explore gene expression with different computational tools. To confirm the importance of a set of genes associated with a phenotype, complementary techniques such as quantitative PCR (qPCR) are mandatory. Thus, arrays are not simply a way to find single differentially regulated genes; they can be used to compare global gene expression in distinct groups.

We show that SpA patients have an aberrant gene expression profile in comparison to healthy controls, indicating that alteration of gene expression in the colon of SpA patients is a biologically relevant concept. We identified a set of genes that are differentially expressed in both CD and SpA patients who are at higher risk of developing CD. Based on the expression of these 95 genes, SpA patients with subclinical chronic gut inflammation cluster with CD,
confirming the clinical association between the two inflammatory disorders. We also suggest a number of candidate genes for mutation screening. We are currently verifying a selection of genes by qPCR, and exploring the involvement of genes that are differentially expressed in both CD and SpA patients with a history of chronic gut inflammation, in order to find early (genetic) markers for CD in SpA patients.

REFERENCES


Demetter, P., J. A. Van Huysse, F. De Keyser, et al. (2002). "Increase in lymphoid follicles and leukocyte adhesion molecules emphasizes a role for the gut in spondyloarthropy


Stoll, M., B. Corneliussen, C. M. Costello, et al. (2004). "Genetic variation in DLG5 is


THE IMPORTANCE OF GENE EXPRESSION PROFILING IN MAPPING COMPLEX GENETIC TRAITS

Debby Laukens

Department of Molecular Biomedical Research, Ghent University and Flanders Interuniversity Institute for Biotechnology (VIB), Gent, Belgium

INTRODUCTION

Microarrays are the ideal tool to study many thousands of genes at once. However, since their widespread use, little has been accomplished, except for simply identifying differentially expressed genes. Usually, the processing and interpretation of the huge amount of data is difficult. Another application based on microarrays is the extraction of new classification algorithms. However, this demands a lot of samples, and is thus not cost-effective. Additionally, microarrays are used for the identification of pathways. However, many genes will be missed, because not all expression changes will occur at one specific time-point. Together with the statistical set backs, these are the reasons why studies on microarrays are increasingly criticized. They seem to be merely applicable as an initial screen for differentially expressed genes, and absolutely require additional verification. Moreover, it was recently postulated that the use of microarrays to study transcription profiling in the future might become less important because “it focuses on an intermediate” (Hoheisel 2006). Indeed, RNA might be too far removed from the actual cellular effectors to be used for diagnostic purposes, unless regulatory RNAs, such as microRNAs are targeted (He et al. 2004). We believe, however, that the RNA expression of genes is crucial in the development of diseases, and is probably a major determinant for common phenotypic characteristics, such as height or intelligence. Therefore, transmitted gene expression patterns associated with complex genetic diseases should provide useful information, but up to now, this has been underestimated in genetic studies. Here, we consider the use of microarrays in the identification of susceptibility genes, and we propose a novel method based on the integration of microarray analysis in family pedigrees.
NATURAL VARIATION AND SUSCEPTIBILITY

Natural genetic variation at the DNA level greatly influences the basal expression of genes or their targets. Analysis of microarrays to investigate the genetic circuitry that regulates expression has first been used in yeast (DeRisi et al. 1997). It is a favourable organism to study gene expression, since the genes are easy to recognize, and the cis regulatory elements are compact and close to the transcription units. Certain yeast mutants have been characterized by their expression pattern and sets of genes could be grouped and rationalized by identification of upstream regulatory sequences. In mice, differential gene expression induced by single-gene effects in complex tissues has been studied (Callow et al. 2000; Aronow et al. 2001). In contrast to the exploration of functional consequences of a defined genetic difference, these approaches have also been used in combination with traditional quantitative trait locus (QTL) mapping techniques. Indeed, Karp and co-workers have successfully identified complement factor 5 (C5) as a susceptibility gene in a murine model for allergic asthma (Karp et al. 2000). C5 appeared to be the only gene that was differentially expressed between susceptible and non-susceptible backcross mice as well as located near one of the previously identified QTL intervals. Similarly, Aitman and colleagues have combined a congenic strain strategy with micorarray expression analysis in the spontaneously hypertensive rat, a model for human insulin-resistance type 2 diabetes (Aitman et al. 1999). Hence, CD36 was pursued as a candidate for the observed QTL at chromosome 4, and sequencing revealed multiple variants caused by unequal genomic recombination of a duplicated ancestral section.

In humans, little has been studied on gene expression and susceptibility. However, remarkable support comes from the study of allele specific variation in gene expression (Lo et al. 2003). Among 602 heterozygous genes that contained transcribed single nucleotide polymorphisms, 54% displayed at least a twofold preferential expression of one allele. This allelic variation was shown to be transmitted by mendelian inheritance (Yan et al. 2002). Furthermore, the variance in expression level is highest among unrelated individuals, and smallest between members of monozygotic twin pairs, suggesting that germ-line differences contribute to variation in gene expression (Cheung et al. 2003; Correa et al. 2004). Four studies illustrated that basal gene expression levels are highly heritable (Watts et al. 2002; Cheung et al. 2003; Correa et al. 2004; Pastinen et al. 2004). Watts and co-workers have defined a unique expression phenotype to carriers of a recessive disease called ataxia telangiectasia. Although many studies have addressed the inheritance of single-gene mutations, so far the genome-wide inheritable expression in human disease has not been studied (Cheung et al. 2002). Moreover, Morley and colleagues
Future prospects

have shown that these quantitative traits allow the genetic mapping of determinants that contribute to variation in human gene expression (Morley et al. 2004). Taken together, these results demonstrate the feasibility and utility of exploring genome-wide gene expression in the study of susceptibility to human diseases.

Polymorphisms and gene expression

How can polymorphisms affect gene expression? First of all, mutations in coding sequences often result in an altered half-life of the RNA. Secondly, polymorphisms in untranslated regions, or introns can have modest or substantial influences on gene expression. Not only transcription factor binding sites or enhancer sites can be changed, but also the stability of the RNA is depending on seemingly “junk” sequences. Next to changes in cis acting elements, polymorphisms in trans acting factors could influence the expression of downstream genes. The small variations in DNA sequences between individuals have been suggested as being responsible for inter-individual phenotypic variation. That is why common traits or phenotypic characteristics are supposed to result from natural variation. In fact, all forms of life, including life-threatening diseases and normal phenotypes, probably share the same basic idea: it is merely the impact of the polymorphism that determines the functional consequence (Figure 1). If the impact is low, such as small phenotypic changes, the polymorphism will not be eradicated by natural selection. On the other hand, if the polymorphism is causing a severe dysfunction of a protein, it will be selected out. That is probably one of the reasons why some relatively harmless traits such as allergies are common.

Epigenetics and gene expression

Unfortunately, the classical genetic ideas are complicated by new and slowly accepted mechanisms. The concept of the all-determining DNA sequence is not complete. It appears that the environment not only affects your own life, but also the life of your progeny. It is becoming widely accepted that genes have some sort of “memory”. The way this memory works is explained by epigenetic mechanisms (literally: on top of genetics), acting by switching on and off the expression of genes. This also gives a new dimension as to what “environmental factors” are in genetics. Are they the factors that the person goes through at birth, during adolescence or perhaps that his ancestors were confronted with? Epigenetic effects can act on larger regions, influencing a number of genes within the same chromosomal region. Moreover, these effects are not fixed over time within the same individual. Today, two basic epigenetic mechanisms are known: DNA methylation and histone modifications. These modifications act on the expression
Future prospect

of genes by interfering with chromatin packaging or the physical binding of *trans* factors to the DNA strand. Furthermore, at least for methylation, evidence exists that it is genetically transmitted.

![Graph showing epigenetic effects, frequency of trait, impact of polymorphisms, number of polymorphisms, and classification of traits into "every-day" phenotypes or quantitative traits, common traits, less common traits](graph.png)

**Figure 1** Natural genetic variation determines every-day life: from life-threatening diseases to the way we look.

**FAMILY-BASED EXPRESSION PROFILING IN CROHN’S DISEASE**

As stated above, susceptibility to CD most likely results from several variations in DNA sequence. Some of them will alter the gene product, but they may also have profound as well as subtle consequences on gene expression. On top, inherited epigenetic mechanisms will contribute to differential expression. Instead of analyzing gene expression in randomly selected CD patients, we suggest using multiple affected families to screen for sets of genes that are differentially expressed, e.g. in blood, between affected and non-affected family members (Figure 2). This way, the problem of genetic heterogeneity between random CD patients is dramatically reduced. Specific sets of genes can be identified that are transmitted only to patients with CD, and are thus ideal candidates to associate to disease. In combination with linkage data of the families, we would gain substantial evidence for an actual expression signature for CD, in which genes will have to be characterized. In the next section we briefly describe the experimental set-up of such a study.
Design and methodologies

Gathering of appropriate families/patients and blood sampling. Patients and relatives will be asked to participate in the study. Families will be evaluated based on the magnitude, number of affected individuals, and genetic information. Patients within these families will be selected only after diagnosis is well defined, and non-affected patients should have reached a certain age, to reduce false negatives.

Collected fresh whole blood will be immediately treated with the RiboPure-Blood Kit (Ambion). This way, samples are safe to transport at ambient temperature. The majority of RNA will originate from white blood cells (WBC). The RiboPure-Blood Kit was designed to minimize contamination with heme, protein and genomic DNA during RNA extraction of whole blood. Therefore, there is no need to select for WBC before extraction, reducing experimental heterogeneity. It has been reported that some features of variation in expression in blood samples are associated with cellular composition, gender, age and time of the day, so care must be taken...
**Future prospect**

in this regard (Whitney et al. 2003). To rule out such random effects, sampling will be done at two time points per individual, and expression profiling will be carried out in duplicate. In addition, variations observed among patients are likely to be significantly greater than the background variation in normal gene expression.

**RNA extraction and microarray screening.** Total RNA will be isolated within 1 day, and stored until it can be processed for RNA amplification and hybridization. We choose to perform a genome-wide screen and use the Affymetrix platform (Human Genome U133 Plus 2.0 Set, containing 47,000 transcripts including 38,500 well-characterized genes). These chips contain oligonucleotide probes, designed to maximize sensitivity, specificity, and reproducibility, allowing consistent discrimination between specific and background signals, and between closely related target sequences.

**Data analysis.** The raw expression data will be imported in a highly developed software program called GeneMaths® XT (Applied Maths, BVBA).

In first instance we will concentrate on those genes that are annotated and modulated between affected and non-affected individuals, base on a simple P-value cut-off. Differentially regulated genes will belong to one of the following main categories:

- false positives (due to limited statistical power)
- secondary to disease (this should be minimized, since we are not looking at the site of pathology, and genes should be at their steady-state level. In addition samples are analyzed at two time-points per individual)
- genetic variation in *trans*-acting genes
- genetic variation in *cis*-acting elements

An example of the concept and analysis of *cis* and *trans* regulators of expression in human genetics has been anticipated (Morley et al. 2004). Genes that are located within regions with most significant evidence of linkage are potentially those that are regulated in *cis*. Based on the information of the genes we will have to decide on which genes to focus. Genes that are not located within these regions can still be regulated in *trans*. Extra evidence will have to be gathered to decide if and which genes might be interesting to pursue. For example, unsupervised clustering can reveal genes that are commonly regulated. Promoter analysis will be
conducted to search for common binding factors. Evidently, transcription factors that are located within a locus are considered as candidate genes.

In a more elaborate search for regulators, the expression phenotype can be analyzed genetically as a quantitative trait in order to identify the determinants of the variation in gene expression. The differentially expressed genes with the most evidence of linkage will be used to carry out genome-wide linkage analysis in CEPH families (Centre d'étude du polymorphisme humain) (Morley et al. 2004). These are large three-generation families, many of which have been genotyped. Immortalized lymphocytes are available from all the members of the pedigree. Consequently, loci can be attributed to specific expression phenotypes, and be evaluated by comparing to known CD loci.

**Genetic screening.** Genes that meet the criteria as discussed above, will be screened for mutations/polymorphisms in matching patients. For this, sequence analysis of exonic sequences and ~1000 nucleotides upstream of the start codon will be performed. Polymorphisms are then evaluated in an independent CD and control population to assess the population risk.

**REFERENCES**


Future prospect


Summary and discussion

The objective of this project was to use gene expression analysis in the search of genetic susceptibility genes for Crohn’s disease (CD). Crohn’s disease is a debilitating disease with substantial impact on the patients’ quality of life. For some reason, the intestine becomes chronically inflamed, extending deep into the layers of the intestinal wall. The patient experiences periods of clinical relapses and remission. In general, the complete gastrointestinal tract can be affected, however, the inflammation is usually confined to the small bowel (ileum) or the large bowel (colon). In fact, the location of disease is one of the most accepted phenotypic classifications for CD, because it stays particularly stable throughout disease duration. Moreover, concordance rates for disease location in affected siblings are high.

The principle treatment for CD is the lifetime controlling of gut inflammation, often resulting in surgical resection of parts of the intestine. However, because of substantial side effects and uncontrollable relapses, conventional treatment remains far from satisfactory for patients and physicians. A more rational approach for developing therapies should evolve from a good knowledge of the pathogenesis: when and why is inflammation initiated, how does it result in a chronic state and why is the patient not able to control the inflammation? A very important question is which genetic or environmental factors are common in most CD patients. It is generally accepted that good hygiene standards contribute significantly to the development of CD. This so-called hygiene-hypothesis is based on the reality that today, people encounter too little infections, resulting in a less trained immune system. Indeed, in developing countries, infectious diseases are fairly common, yet CD does practically not exist. Nevertheless, it is still a guess whether CD arises from the loss of tolerance to commensal bacteria or whether it is infection-triggered. In any case, the role of intestinal flora in genetically predisposed people is undisputable. Today, most studies are directed at this hypothesis.

In 2001, two independent research groups mapped the first susceptibility gene for CD, CARD15. This gene is located at the IBD1 locus (16q12), a locus that was identified in almost every genome screen. The CARD15 protein is a cytosolic sensor for bacterial components. Up to one-third of CD patients have a probable loss-of-function mutation in this gene. The discovery of CARD15 and CD-associated mutations was a major advance in unravelling some of the characteristics of CD pathogenesis. Firstly, it explains a great deal about the involvement of bacteria in the disease. For example, it was shown that epithelial cells overexpressing a mutated
CARD15 protein were not able to respond to infection with *Salmonella*. Secondly, these mutations are associated with ileal involvement. Interestingly, CARD15 is expressed in the paneth cells, which are specialized cells that secrete anti-bacterial molecules into the lumen of the gut. These paneth cells are normally restricted to the ileum, which fits with the phenotype of ileal affection in CD.

However, maybe *CARD15* should not be considered as being mutated in the strict meaning of the word (see also *Intermezzo*). In up to 20% of healthy persons, these mutations are also present. A gene like *CARD15* is most likely not a causative gene for CD, but rather a disease-modifying gene. Therefore, we better consider them as polymorphisms. An interesting hypothesis stated by Hugot and co-workers fits well within this concept. They speculated that these polymorphisms, which arose quite recently, were somehow beneficial during the outbreak of the plague in Europe. Later, the number of CD patients increased, partly because of increased hygiene standards. Because CARD15 is predominantly expressed in the ileum, patients carrying the “altered” CARD15 protein show more inflammation of the ileum.

The example of *CARD15* and the implications it had on research is substantial: a pubmed search on CARD15 gives 501 hits since 2001. Thus, the search for new functionally relevant proteins in the form of susceptibility genes is essential for a thorough understanding of mucosal biology in general and pathogenesis of CD in particular. The classical genetic linkage studies on CD have been crucial in the identification of *CARD15*. Therefore, we reviewed all the data of genetic linkage and gene association studies in Chapter 1. As for most complex genetic traits, many so-called chromosomal disease loci have been identified. Unfortunately, the exact genes within these loci that are associated with CD are not known. In contrast to simple mendelian traits, finding a master gene for the causation of CD is not plausible. In reality, the combination of many subtle polymorphisms with environmental risk factors and epigenetic mechanisms will most likely be responsible for disease onset. Another method for finding susceptibility genes is the candidate gene approach. Here, a gene is chosen based on a known property. Subsequently it is screened for mutations in several CD patients, and frequencies are compared in a larger case-control study. Though many genes have been screened in this manner, the reproducibility of association studies is disappointing. Mainly, this is the result of statistic setbacks and population biases. We gave an overview of all genes that were tested in such association studies. Two main conclusions can be drawn from this review. Firstly, patient and control cohorts should be large and thoroughly chosen. Often, populations are too small, and control cohorts are not age and sex matched. Good collaboration and communication of research labs with clinicians and hospitals is crucial. For most small laboratories, unfortunately, the use of undersized populations for
Summary and discussion

"every-day" phenotypes or quantitative traits (intelligence, height)  common traits (allergies, hypertension)  less common traits (cystic fibrosis, mongolism) or rare phenotypes (albinism)

penetrance problems: are they the result of additional polymorphisms in protecting or sensitizing genes?

Intermezzo: Natural genetic variation determines every-day life: from life-threatening diseases to the way we look.

Any two humans are approximately 99.9% identical in their DNA sequences. However, the 0.1% variation, that is one dissimilarity per 1,000 bases on average, contributes to difference in the risk of developing certain diseases, and responses to drugs, infectious agents, toxins and other environmental factors. Finding the genetic variants that influence disease risk and drug response is necessary to understand how genetic and environmental factors interact to influence health. In fact, the foundation of all forms of life, including life-threatening diseases and normal phenotypes, probably share the same basic idea: it is merely the functional impact of the polymorphism that determines its frequency. If the impact is low, such as small phenotypic changes, the polymorphism will not be eradicated by natural selection. On the other hand, if the polymorphism is causing a severe dysfunction of a protein, it will be selected out. That is probably one of the reasons why some relatively harmless traits such as allergies are so common. Incomplete disease penetrance is a term that is used to describe individuals with a specific genotype who do not manifest that genotype at the phenotype level. In reality, this might reflect the absence or presence of polymorphisms in other sensitizing or protecting genes respectively. In addition, epigenetic mechanisms complicate this generalized mechanism. They influence the expression of genes, and represent thus another level of determining penetrance. Probably, but this has not been proven, they will not be important in simple mendelian traits. Moreover, epigenetic influences also give a new dimension as to what “environmental factors” are in genetics. Are they the factors that the person goes through at birth, during adolescence or perhaps that his ancestors were confronted with?
association studies is the only means of competing with the more organized labs. Nevertheless, even small scale pooling of cohorts can improve the outcome of genetic screenings. In 1997, a group of scientists and clinicians founded the IBD International Genetic Consortium (IBDIGC), now a group of twelve research groups involved in studying genes that are implicated in IBD. They collaboratively study large numbers of well-documented families for linkage. The formation of large international consortia offers research groups both large and small to participate equally in CD gene identification. The second conclusion is that more standardized methodologies for the screening of genes are required. In addition, we should evolve from single polymorphism association studies to the association of so-called haplotype structures. These haplotype blocks are the minimal genomic fragments that are genetically transmitted. A single polymorphism within a gene might not necessarily influence its function or transcription. Yet this defect could be caused by a nearby polymorphism within the same haplotype block. For instance, the CARD15 gene contains, next to the CD-associated mutations, many other polymorphisms dispersed all over the gene. Because not all linkage observed at 16q12 can be explained by the three CD-associated polymorphisms, the question arises whether the actual disease causing mutation is located near the CARD15 gene. In Chapter 2, we conducted a search for polymorphisms in the CARD15 promoter region, because the expression of CARD15 is up-regulated during inflammation. We found a common polymorphism, c.-59G>A, located within a DNA stretch that potentially binds the E2F transcription factor. Using an in vitro reporter system, we showed that the promoter is less responsive to tumor necrosis factor alpha induction in the presence of the c.-59A allele. This functional defect prompted us to study this polymorphism in more detail. It appears that the -59A allele is always linked to the three other CD-associated mutations. This allele occurs isolated, but it is always present whenever one of the three CD-associated mutations occurs. Therefore, we decided to study the haplotype structure surrounding CARD15, which is currently ongoing.

The selection for a gene to study in a candidate gene approach might require some adjustment. The choice of a gene is very subjective, and virtually every gene can theoretically fit somewhere within the complexity of CD pathogenesis. We used a complementary and hypothesis-independent approach to identify more reliable candidate genes for CD in Chapter 3. A gene expression survey on unaffected colon biopsies from CD patients and healthy controls was used to identify genes that are differentially expressed in CD. We then focused on those genes that are located within an arbitrary region of 5 centimorgan from a marker that showed highest linkage to CD in at least two independent studies and in a study performed on a Flemish cohort. Eighteen genes were selected in this way. These genes can individually serve as new
focus genes, also for other researchers within the field. We focused on two genes within this list that belong to a cluster of highly related genes called metallothioneins (MT). These genes are located in tandem within the IBD1 locus. Moreover, its role in host protection against various forms of stress and inflammation made them ideal candidate genes to pursue. We showed that their expression is significantly down-regulated in colon, ileum and blood of CD patients that have colonic disease. In addition, a reduced protein expression in this type of patients was shown by immunohistochemistry. Remarkably, we showed a correlation between RNA and protein levels, even though biopsies used for RNA extraction and for staining were not taken at the same moment. This further supports the idea that the altered MT expression is genetically determined.

We found no coding or promoter mutations in one of the MT genes, MT1M. Therefore, we screened a well-characterized MT transcription factor, MRE-binding transcription factor 1, MTF1, for mutations. Two missense mutations were found in 6 out of 95 patients. Their functional relevance still needs to be determined. A frequent polymorphism in the splice site junctions at the boundary of exon 8 was found, however, we could not detect any splice variants.

In our study, we focused on a polymorphism in the first intron, IVS-128A>T, which is a potential binding site for the GATA4 transcription factor. Although we showed that the frequency of this polymorphism was not different in CD patients as compared to healthy controls, it was significantly associated with disease location. The apparent resemblance to the association of CARD15 with disease location encouraged us to study the combined risk for developing ileal disease. The odds for developing ileal disease was 4 in CARD15 mutant patients, and 2.4 in patients that carry the IVS-128T allele. Moreover, we showed a gene-gene interaction of CARD15 and MTF1 as shown by a higher combined risk. Theoretically, this means that when CD patients are genotyped for CARD15 and MTF1, ileal involvement (or alternatively, pure colonic disease) during disease progression can be predicted with substantial precision. Moreover, this is to our knowledge the first report of the use of microarray screening of tissues in genetic studies. We described the importance of gene expression profiling for the mapping of new susceptibility genes in the Future prospects, and provided a novel method based on transmission of expression signatures in family studies.

The involvement of MTF1 in the genetic down-regulation of MTs in CD patients needs to be determined. Interestingly, both CARD15 and MT are highly expressed in the ileum, and are involved in bacterial detection and eradication respectively. Bacterial load in the ileum is $10^{2-3}$ per gram faeces, while it reaches $10^{9-12}$ in the colon. It thus seems likely that the ileum is specialized in eliminating bacteria from its environment. CARD15 and MT may belong among the players that regulate this elimination. Conversely, the observation that mutant CARD15 and
Summary and discussion

the IVS-128T allele predispose to ileal disease could even so be reversed. Possibly, a CARD15 mutated protein or the MTF1 protein in IVS-128T allele carriers, protects against an unwanted colonic inflammation, which might occur as a result of the high bacterial load at the colonic site. Importantly, the next priority is examining MT expression and assessing the risk of MTF1 and CARD15 in ulcerative colitis (UC) development, which is another related inflammatory bowel disease. Indeed, UC pathology could be considered as the extreme end of the disease location phenotype of CD. In UC patients, only the colon is affected but inflammation is disturbing only the superficial layers of the bowel wall. Although UC and CD are fundamentally different diseases, the concept of disease location might be influenced by the same molecular mechanisms.

We and others have demonstrated that MTs are highly expressed in intestinal epithelial cells (IEC). Over the past decade, many studies have revealed the physical and immunological importance of IEC in the gut homeostasis. These cells have developed a variety of mechanisms to reduce the risk of infection or damage by toxic compounds or free radicals. The effective maintenance of a physical barrier function is dependent on the establishment of well-organised intercellular junctions and a constant state of regeneration/renewal of the epithelium. Intestinal epithelial cells also participate in the innate immune responsiveness of the intestine by their ability to secrete mucus, antimicrobial peptides, cytokines and as well as chemokines. We generated an MT-knockdown model of HT29 colonic epithelial cells to study the biological consequences of a reduced MT expression in IEC. Initial experiments revealed that in comparison to the wild type cell line, these cells produce less IL8 when they were challenged invasive E. coli. This might provide a direct pathogenic role, since CD was previously linked to immunodeficiency caused by impaired innate immunity. In addition, preliminary data showed that these cells do not induce IL8 secretion in response to commensal bacteria (Lactobacillus rhamnosus), therefore an infection model for CD may fit best with our hypothesis. Nevertheless, a number of other bacterial strains, both commensal and pathogenic, need to be tested in this system.

Metallothioneins have been linked to a number of diseases, including some neurological disorders and different types of cancers. However, the regulation of MT isoforms in human cells is very complex, and its expression in health and disease is difficult to interpret. An overview of MT involvement in diseases, together with the regulation of MT expression is given in Chapter 4. For CD, conflicting data on MT expression are found in literature. The same holds true for MT expression in many cancers. The main problem in comparing these studies is the use of different techniques used for quantification. Since the original cloning of the first MT gene from
the renal cortex in 1960, original data on MT quantification in serum or on histological sections greatly depend on the antibody used. It is obvious that cross-reactivity between species but also between isoforms complicates the interpretation of results. These days, many researchers use the same anti-MT antibody, which facilitates comparing results. Recently, more detailed information on the expression on the different MT isoforms is gathered because of the emergence of more sensitive and specific techniques such as quantitative real-time PCR. In order to learn why MTs are not adequately expressed in CD, we made a literature overview of human MT regulation. This overview was based on an *in silico* comparison with murine expression control, because the majority of studies on MT were performed in mice. The genes upstream of MT are good candidates to study with respect to MT expression in CD.

In the next section, we used gene expression analysis to study the well-defined association between CD and another group of chronic inflammatory diseases called spondyloarthropathies (SpA). In addition, the genetic contribution of CD-associated *CARD15* mutations was assessed in these patients. SpA covers a heterogeneous group of arthritic diseases sharing clinical and radiological features and presence of the *HLA-B27* allele. Previous research has indicated that patients with spondyloarthritis (SpA) form a population with a substantial higher risk for the development of CD, which is discussed in Chapter 5. Subclinical inflammatory gut lesions are common in SpA, as observed during colonoscopic studies in these patients. These gut lesions are either classified as acute or chronic inflammation. The latter form resembles closely lesions found in CD patients. Follow-up studies of SpA patients indicated that 7% develop an inflammatory bowel disease and, of those with chronic inflammatory gut lesions, 30% develop clinical CD. Several lines of evidence suggest that the gut could have an important pathogenic role in SpA. However, the precise relation of the join-gut axis is not completely understood. We therefore assessed the contribution of *CARD15* mutations in SpA patients, as described in Chapter 6. We found that the frequency of *CARD15* mutations in SpA patients with chronic gut lesions was significantly higher than in the control population, and comparable to the frequency in CD patients. Thus, we provided a first line of evidence that gut inflammation in SpA patients is genetically determined, and that SpA patients with subclinical chronic gut inflammation are a separate group, with a higher risk to develop CD.

We wondered whether the intestinal gene expression pattern in SpA patients differs from that of controls. In Chapter 7 we described alterations in gene expression in unaffected colon biopsies of SpA patients as compared to controls. Furthermore, the alterations found in patients with a history of chronic gut inflammation are comparable to those found in CD patients,
Summary and discussion

confirming the clinical observations. Hence, SpA serves as a good model to study early CD. Furthermore, the SpA model might be used in the selection for potential candidate genes for the predisposition of gut inflammation.
Samenvatting en discussie

De opzet van dit project bestond erin genexpressie te bestuderen om genetische afwijkingen voor de ziekte van Crohn (ZVC) op te sporen en te karakteriseren. Patiënten met de ZVC lijden aan een chronische ontsteking in de darmen. Hierbij is de volledige darmwand, zowel de oppervlakkige als de diepere lagen, aangetast. Het ziektebeeld is sterk individueel bepaald en kent een grillig verloop: periodes van hevige klachten en kalmere periodes wisselen elkaar af. De belangrijkste verschijnselen van de ZVC zijn een rechtstreeks gevolg van de ontsteking zelf, zoals pijn in de buik, koorts, krampen, al dan niet bloederige diarree en gewichtsverlies. Vooral het uiteinde van de dikke darm (colon) en de dunne darm (ileum), maar ook andere delen van het maag-darmstelsel kunnen worden aangetast. Het is zelfs zo dat de plaats van ontsteking wordt gebruikt om patiënten met de ZVC te classificeren, omdat deze relatief stabiel blijft tijdens het ziekteverloop. Bovendien is er een sterke overeenkomst in de plaats van inflammatie tussen verwanten met de ZVC.

De ZVC kan niet worden genezen, dus is de behandeling erop gericht de ontsteking onder controle te houden en complicaties te verminderen. Vaak moeten patiënten na verloop van tijd een chirurgische ingreep ondergaan, gaande van het weghalen van een abces tot het verwijderen van een deel van de zieke darm of zelfs het volledige colon. Hoewel veel patiënten aanvankelijk geholpen zijn met de standaardbehandelingen treden er vaak neveneffecten op, en kunnen perioden van herval niet worden gecontroleerd. Een meer rationale aanpak gebaseerd op kennis van het ziekteproces is noodzakelijk bij het ontwikkelen van nieuwe behandelingen. Hierbij moeten we enkele concepten proberen te begrijpen: wanneer en hoe wordt de ziekte geïnitieerd, waarom evolueert een ontsteking tot chronisch lijden, en waarom is de patiënt niet in staat de inflammatie onder controle te houden? Belangrijk is ook de vraag welke genetische en omgevingsfactoren patiënten gemeenschappelijk hebben. Een algemeen aanvaarde hypothese is de zogenaamde hygiëne-hypothese. Ze gaat ervan uit dat we tegenwoordig te weinig infectieziekten doormaken, waardoor het afweersysteem onvoldoende geoefend wordt. Inderdaad, het is opmerkelijk dat er bijna geen ZVC voorkomt in ontwikkelingslanden, waar er echter veel infectiegevaar dreigt. Nochtans weten we nog altijd niet of patiënten die de ziekte ontwikkelen een verminderde tolerantie vertonen in de darm, of dat de ziekte een initiële infectie vereist. Hoe dan ook, we gaan er van uit dat de darmflora een belangrijke rol speelt in de ZVC, maar enkel bij personen met een genetische aanleg de ziekte uitlokt. Dit is dan ook vandaag de dag het onderwerp van veel onderzoek in de ZVC.
Het eerste gen dat sterk werd geassocieerd met de ZVC werd gekloneerd in 2001. Dit gen, *CARD15* genaamd, ligt in de *IBD1* locus (16q12), een locus die in bijna elke genoomscan werd geïdentificeerd. Het CARD15 eiwit herkent bacteriële componenten binnen in de cel, en activeert NFκB. Ongeveer een derde van de patiënten met de ZVC dragen een mutatie in dit gen. Veel onderzoek was aanvankelijk gebaseerd op de functie van deze mutaties, maar hierover bestaat nog geen sluitend antwoord. Er zijn echter sterke aanwijzingen dat het gemuteerde CARD15 eiwit zijn functie gedeeltelijk niet meer kan uitoefenen. Over het algemeen was de ontdekking van *CARD15* van groot belang om enkele aspecten in het ziekteverloop van de ZVC te begrijpen. Ten eerste bieden ze een verklaring voor het belang van bacteriën in het ziekteproces. Een mooi voorbeeld hiervan is dat epitheliale cellen die het gemuteerde CARD15 eiwit overexpresseren niet meer reageren op infectie met *Salmonella*. Ten tweede, mutaties in *CARD15* werden gecorreleerd met ileale aantasting van de ZVC, de plaats waar CARD15 sterk tot expressie komt. CARD15 wordt geëxprimeerd in paneth cellen, gespecialiseerde cellen die bacteriën in het ileum elimineren. In normale omstandigheden komen deze cellen enkel voor in het ileum, wat een directe link vormt tussen de werking van CARD15 en de plaats van aantasting.

Misschien is het belangrijk om de mutaties in *CARD15* eerder te beschouwen als polymorfismen die een invloed uitoefenen op het ziekteproces (zie ook *Intermezzo*). Ze komen ook voor bij 20% van gezonde personen. Het is geen oorzakelijk gen, maar eerder een gen dat een invloed heeft op de manifestatie van de ziekte, in dit geval, ileale aantasting. Deze redenering sluit goed aan bij een hypothese die Hugot en medewerkers vooropstelden. Zij geloven dat personen die deze polymorfismen in *CARD15* droegen, bevoordeeld waren bij het uitbreken van de pest in Europa. De oorsprong van deze polymorfismen loopt inderdaad gelijk met deze periode. Vanaf dan is de incidentie van de ZVC beginnen stijgen, mede door een hogere hygiënestandaard. Personen met mutaties in *CARD15* en de ZVC ontwikkelen voornamelijk ontsteking in het ileum, waar CARD15 voornamelijk tot expressie komt.

De impact van *CARD15* op het onderzoek naar de ZVC, maar ook naar aangeboren immunititeit is groot: het resultaat voor CARD15 in pubmed geeft 501 hits sinds 2001. Dit betekent dat de identificatie van nieuwe eiwitten betrokken bij genetische predispositie, van belang is bij de studie van mucosale biologie en van de ZVC in het bijzonder. De klassieke genetische *linkage* studies waren cruciaal bij het klonen van *CARD15*. Daarom hebben we een overzicht gemaakt van alle *linkage* en associatie studies bij de ZVC in *Chapter 1*. Zoals voor de meeste complex genetische aandoeningen werden er meerdere chromosomale loci gevonden voor de ZVC. De precieze genen binnen die loci die geassocieerd zijn met ZVC werden echter
Samenvatting en discussie

Epigenetische effecten

Frequentie van de aandoening

Impact van het polymorfisme

Aantal polymorfismen

“Alledaagse” fenotypes of quantitatieve fenotypes (intelligentie, grootte)

Vaak voorkomende ziekten (allergieën, hypertentie)

Zeldzame ziekten (cystische fibrose, mongolisme) of fenotypes (albinisme)

Penetrantie: is dit het resultaat van additionele polymorfismen in modificerende genen?

Intermezzo: Natuurlijke genetische variatie bepaalt het leven van alledag, gaande van levensbedreigende ziekten tot hoe we er uit zien.

Het genoom van twee individuen is 99.9% identiek. De 0.1% variatie in DNA sequentie is echter bepalend voor de individuele gevoeligheid voor bepaalde aandoeningen of de respons op geneesmiddelen, toxische stoffen en andere omgevingsfactoren. Indien de genetische varianten voor dergelijke gevoeligheden kunnen worden geïdentificeerd, kunnen we beter begrijpen hoe bepaalde ziekten tot stand komen. Als we de lijn verder doortrekken kunnen we aannemen dat alle levensvormen, gaande van levensbedreigende ziekten tot normale fenotypische karakteristieken, draaien rond hetzelfde basisidee: de functionele consequentie van een polymorfisme bepaalt in grote mate hoe vaak het polymorfisme voorkomt. Als de impact klein is, zoals bescheiden fenotypische veranderingen, zal het polymorfisme niet verdwijnen door natuurlijke selectie. Als de impact van een polymorfisme daarentegen groot is en bijvoorbeeld een ernstige verandering van een cruciaal eiwit veroorzaakt, zal het worden uitgeselecteerd. Dit is waarschijnlijk een reden waarom sommige relatief onschuldige ziekten zoals allergieën vaak voorkomen. Onvolledige ziekte-penetrantie is een term die wordt gebruikt om te beschrijven dat iemand een bepaald genotype draagt, maar het fenotype niet tot uiting brengt. Dit kan mogelijk verklaard worden door het voorkomen van variaties in modificerende genen. Epigenetische mechanismen compliceren dit algemeen schema. Ze beïnvloeden de expressie van genen, en kunnen op die manier dan weer beschouwd worden als bepalend voor penetrantie. Wellicht, maar dit werd niet bewezen, spelen epigenetische mechanismen geen of weinig rol bij mendeliede ziekten. Epigenetische invloeden geven ook een heel nieuwe kijk op wat omgevingsfactoren zijn voor de klassieke genetica. Zijn dit factoren die een iemand meemaakt tijdens de geboorte, gedurende adolescentie, of zijn dit misschien factoren die zijn voorouders hebben doorgemaakt?
nog niet geïdentificeerd. In tegenstelling tot mendeliaanse genetische ziekten verwacht men geen geïsoleerd dominant gen dat de oorzaak is bij alle patiënten met de ZVC. Wellicht zal de combinatie van vele polymorfismen, omgevingsfactoren en epigenetische mechanismen aan de basis liggen van de ZVC. Een veel gebruikte alternatieve manier om genen op te sporen is de zogenaamde kandidaatgen benadering. Hierbij wordt er een arbitrair gen gekozen dat mogelijk betrokken is in de ziekte, en zoekt men bij een klein aantal patiënten naar mutaties in dat gen. Vervolgens wordt de frequentie van eventuele mutaties vergeleken bij een groot aantal patiënten en controles. Veel genen en mutaties werden op deze manier geassocieerd met de ZVC, maar de reproduceerbaarheid is telkens minimaal. Dit is voornamelijk het gevolg van beperkingen in statistische methoden en populatie bias. In Chapter 1 geven we een overzicht van alle genen die werden getest op associatie. Hieruit kunnen we twee belangrijke conclusies trekken. Ten eerste moet de keuze van de testpopulaties goed overwogen worden, en moet deze voldoende groot zijn. Vaak zijn de populaties te klein, en zijn controles niet gelinkt aan leeftijd en geslacht met de patiëntenzorg. Een goede samenwerking tussen onderzoekslaboratoria, artsen en hospitalen is cruciaal. Zelfs het poolen van cohorten op kleine schaal kan grote impact hebben op de uitkomst van genetische studies. In 1997 stichtten enkele wetenschappers en artsen het *IBD International Genetic Consortium* (IBDIGC). Vandaag telt die groep twaalf onderzoeksgroepen die werken rond genetica van inflammatoire darmaandoeningen. Dit consortium biedt ook kleine groepen de mogelijkheid om op gelijke basis deel te nemen aan de identificatie van genen voor de ZVC. De tweede conclusie die kan worden getrokken uit de vele associatiestudies is dat er meer gestandaardiseerde methodologieën moeten gebruikt worden bij het screenen van genen. Daarenboven zouden we moeten evolueren van de analyse van geïsoleerde polymorfismen naar zogenaamde haplotype blokken. Deze haplotype structuren zijn de minimale DNA fragmenten die worden overgeërfd. Het is best mogelijk dat een geïsoleerd polymorfisme in een gen geen effect heeft op zijn expressie of de functie van zijn eiwit, maar dat een mutatie verderop, binnen hetzelfde haplotype blok, wel een effect heeft. Een mooi voorbeeld hiervan is dat er binnen het *CARD15* gen vele polymorfismen werden geïdentificeerd, waarvan er drie werden geassocieerd met de ZVC. Men kan zich nu afvragen of een andere mutatie binnen het haplotype blok dat *CARD15* bevat, verantwoordelijk is voor een gewijzigde functie van CARD15. In Chapter 2 hebben we gezocht naar polymorfismen in de promoter regio van *CARD15*, mede omdat de expressie van CARD15 opgereguleerd wordt bij inflammatie. We beschrijven een polymorfisme, c.-59G>A, in een regio dat mogelijk de E2F transcriptiefactor bindt. In een in vitro reportersysteem toonden we aan dat het c.-59A allel minder responsief was op tumour necrosis factor alpha inducitie. Dit heeft ons ertoe aangezet om dit polymorfisme meer in detail te
Samenvatting en discussie

bestuderen. Dit polymorfisme komt altijd samen voor met de andere mutaties in CARD15, maar het kan ook geïsoleerd voorkomen. Daarom zijn we gestart met het karakteriseren van het haplotype blok dat CARD15 omvat, om na te gaan of het c.-59G>A polymorfisme deel uitmaakt van een haplotype blok dat geassocieerd kan worden met de ZVC. Dit werk is nu gaande.

Er is nood aan additionele aanwijzingen bij de keuze van genen om te testen in associatie studies. Inderdaad, de keuze van een gen is vaak subjectief, en elk gen kan bij wijze van spreken functioneel gekoppeld worden aan de ZVC. Daarom gebruikten we in Chapter 3 een alternatieve en hypothese-onafhankelijke aanpak om nieuwe en meer betrouwbare kandidaatgenen te identificeren voor de ZVC. In eerste instantie werd een genexpressie-analyse uitgevoerd in gezonde colon biopten. Er werden differentiële geëxprimeerde genen geïdentificeerd tussen patiënten met de ZVC en gezonde controles. Vervolgens selecteerden we genen die binnen een range van 5 centimorgan rond een merker lagen. De merkers die we hiervoor gebruikten lagen in een locus die in ten minste twee onafhankelijke studies werd geïdentificeerd, of in loci die werden geïdentificeerd in een studie uitgevoerd met een vlaamse cohorte. Op deze manier selecteerden we 18 genen. Elk van deze genen kunnen dienen als focus genen voor eenieder die geïnteresseerd is in genetica van de ZVC. Wij hebben ons toegelegd op het karakteriseren van twee genen binnen deze lijst van 18, de metallothioneines (MT). Deze twee genen behoren tot een grote genfamilie die allen in tandem gelegen zijn binnen de IBD1 locus. Een tweede reden om deze genen te bestuderen is dat ze een belangrijke rol spelen bij de verdediging van het organisme tegen verschillende vormen van stress en inflammatie. We toonden aan dat deze genen minder tot expressie komen in het colon, ileum en bloed van patiënten met de ZVC met colon aantasting. Dit werd bevestigd op eiwitniveau. Een interessante bevinding was dat het RNA- en eiwitniveau gecorreleerd waren binnen dezelfde patiënten, hoewel het weefsel voor beide technieken niet op dezelfde dag werd genomen. Dit en de algemeen verminderde expressie in bloed toonde aan dat er een mogelijk genetisch defect is bij de expressie van MT in patiënten met de ZVC die colon aantasting hebben. We vonden geen mutaties in coderende sequenties of in de promoter regio van één van de MT genen, MT1M. Omdat de verminderde expressie van meerdere MT isovormen werd aangetoond, besloten we om de best gekarakteriseerde transcriptiefactor van MT, MRE-binding transcription factor 1 of MTF1, te screenen op mutaties. We vonden twee mutaties die een aminozuurverandering teweeg brachten in 6 van de 95 patiënten met de ZVC. De functionele relevantie hiervan moet echter nog worden bestudeerd. Verder beschreven we een polymorfisme in een splice acceptor positie, maar we hebben geen alternatieve splice variant kunnen aantonen. Wij hebben ons toegelegd op een polymorfisme in het eerste intron van MTF1, IVS-128A>T, omdat deze regio mogelijk de transcriptiefactor
GATA4 kan binden. Er was geen verschil in frequentie van dit polymorfisme in controles en patiënten, maar het heeft net zoals \textit{CARD15} een grote invloed op de locatie van de ziekte. Het risico om ileale aantasting te krijgen is bijzonder hoog als het IVS-128T allel en een van de \textit{CARD15} mutaties aanwezig zijn: de odds ratio is 4 in patiënten met een \textit{CARD15} mutatie, en 2.4 in patiënten met het IVS-128T allel. We toonden ook aan dat er een gen-gen interactie is tussen \textit{CARD15} en \textit{MTF1}, aangezien er een hoger gecombineerd risico is dan verwacht in aanwezigheid van beide factoren. In theorie kan dit betekenen dat we door genotyping van patiënten met de ZVC goed kunnen voorspellen of ze in de loop van de ziekteproces ileale inflammatie, of zuivere colon aantasting zullen ontwikkelen. Tot slot, deze studie is de eerste waar genexpressie wordt gebruikt bij genetische studies. We benadrukken het belang van het bestuderen van genexpressie in predispositie voor complexe ziekten in de \textit{Future prospects}, en beschrijven een nieuwe methode om kandidaatgenen te identificeren aan de hand van de overerving van expressiepatronen binnen families.

De manier waarop MTF1 een invloed kan hebben op de expressie van MT in de ZVC vereist nadere studie. Zowel CARD15 als MT worden voornamelijk geëxpreeserd in het ileum, en zijn betrokken bij respectievelijk de herkenning en eliminatie van bacteriën. Er is een groot verschil in aantal en soort bacteriën tussen het ileum en het colon: in het ileum bedraagt het aantal bacteriën ongeveer $10^{2-3}$ per gram faeces, terwijl dat in het colon oploopt tot $10^{9-12}$. Men kan dus stellen dat het ileum een gespecialiseerde omgeving is waar het aantal bacteriën gelimiteerd moet worden. CARD15 en MT zouden in dit proces een rol kunnen spelen. De observatie dat \textit{CARD15} mutaties en het IVS-128T allel geassocieerd zijn met ileale aantasting kan ook anders worden geënterpreteerd: \textit{CARD15} mutanten en dragers van het IVS-128T allel kunnen beschermd zijn tegen zuivere colon aantasting, omdat de mogelijkheid om een inflammatoire respons op te wekken verminderd is. De hoogste prioriteit is nu om de MT expressie en het IVS-128A>T polymorfisme in combinatie met \textit{CARD15} mutaties te bepalen in patiënten met \textit{colitis ulcerosa} (CU), een andere vorm van inflammatoire darmlijden. Deze ziekte kan tot op zekere hoogte beschouwd worden als het extreme uiteinde van een continu ziektebeeld van ZVC. Bij patiënten met CU is enkel het colon ontstoken, maar alleen de oppervlakkige lagen van de darmwand zijn aangetast. Hoewel CU en de ZVC fundamenteel verschillende ziekten zijn, is het niet ondenkbaar dat er gemeenschappelijke erfelijke factoren bestaan die een rol spelen bij het tot stand komen van inflammatie op een specifieke plaats in de darm.

We toonden aan dat MTs voornamelijk geëxpreeserd worden in epitheliale cellen, een eigenschap die al eerder werd beschreven. De laatste jaren werd het belang van intestinale
epitheelcellen in immunologie aangetoond. Deze cellen bevatten een aantal mechanismen om de kans op bacteriële infectie of schade door toxische stoffen te reduceren. Ze zijn ook betrokken in aangeboren immuniteit: ze produceren anti-bacteriële stoffen en mucus, maar ook cytokines en chemokines. We hebben HT29 colon-epitheelcellen gegenereerd die MT verminderd tot expressie brengen, d.m.v. small interfering RNA. We toonden aan dat deze cellen een verminderde IL8-secretie vertonen in vergelijking met de normale HT29 cellen wanneer ze geïnfecteerd werden met invasieve E. coli. Deze observatie is een eerste directe link tussen een verlaagde MT expressie en de ZVC: deze wordt immers gekarakteriseerd door immunodeficiëntie als gevolg van een gestoorde innate immuniteit. Preliminaire resultaten toonden aan dat deze cellen niet reageren op commensale bacteriën (Lactobacillus rhamnosus). Dit betekent dat een infectiemodel eerder aanleunt bij onze bevindingen. Om hierover zekerheid te krijgen moeten we echter een reeks bacteriestammen testen, zowel pathogene als niet-pathogene stammen.

Metallothioneines werden gelinkt aan een aantal ziekten, waaronder enkele neurologische aandoeningen en kankers. De expressie van humane MT isovormen is echter zeer complex, en de interpretatie van de resultaten is vaak erg moeilijk en onderling niet te vergelijken. In Chapter 4 geven we een overzicht van de expressie van MT in verschillende ziekten, alsook de regulatie van MT in humane cellen. Er bestaan een aantal tegenstrijdige resultaten in verband met MT en de ZVC. Hetzelfde probleem geldt voor MT en kankers. De voornaamste reden hiervoor is dat er verschillende technieken werden gebruikt voor het kwantificeren van MT. Sinds het eerste MT gen werd gecloneerd in 1960, verschenen er veel studies die MT maten in serum en op histologie aan de hand van antilichamen. Cross-reactiviteit tussen species, maar ook tussen verschillende isovormen maakt het moeilijk om deze data te vergelijken. Vandaag wordt echter meestal hetzelfde anti-MT antilichaam gebruikt. Daarenboven maken meer gespecialiseerde technieken, zoals kwantitatieve PCR, het mogelijk om gedetailleerde informatie te krijgen over de expressie van verschillende isovormen. Om een breed beeld te krijgen over hoe een verlaagde MT expressie kan verklaard worden in de ZVC maakten we een overzicht van de regulatie van MT in humane systemen. Dit overzicht is gebaseerd op een in silico vergelijking tussen humane en muis genregulatie, aangezien de meerderheid van de studies gebeurd is op muizen. De genen die aan de basis liggen van MT regulatie zijn uiteraard kandidaatgenen om te testen in de ZVC.

In het tweede gedeelte van het project hebben we gebruik gemaakt van genexpressie om het verband te bestuderen tussen de ZVC en een andere chronisch inflammatoire aandoening,
spondyloartropathie (SpA). Daarenboven werden de *CARD15* mutaties bestudeerd in deze groep van patiënten. SpA is een heterogene groep van verwante aandoeningen gekarakteriseerd door specifieke radiologische en klinische aspecten. Ze vertonen ook een sterke associatie met het *HLA-B27* alleel. Er zijn veel aanwijzingen, beschreven in Chapter 5, dat SpA patiënten een risicogroep vormen voor het ontwikkelen van de ZVC. Darmlaesies zijn een vaak voorkomend fenomeen bij SpA patiënten, zij het in subklinische vorm. Deze laesies kunnen acuut of chronisch van aard zijn. De chronische laesies lijken sterk op de laesies die gevonden worden bij patiënten met de ZVC. Een studie toonde aan dat 7% van de SpA patiënten de ZVC of CU ontwikkelden, en meer specifiek 30% van de SpA patiënten met chronische darminflammatie ontwikkelden de ZVC. Een aantal argumenten wijzen op het belang van deze subklinische darminflammatie bij het ontstaan van SpA. Hoe dan ook, het precieze verband tussen het gewricht en de darm is nog niet volledig begrepen. In Chapter 6 beschrijven we een studie waarbij we de ZVC-geassocieerde *CARD15* mutaties in SpA patiënten beschouwen. We toonden aan dat de frequentie van *CARD15* mutaties in SpA patiënten met subklinische chronische darminflammatie was verhoogd in vergelijking met controles, en vergelijkbaar met de frequentie in patiënten met de ZVC. Met deze studie toonden we aan dat darminflammatie in SpA mede genetisch bepaald is. SpA patiënten met chronische darminflammatie zijn inderdaad te beschouwen als een aparte groep die een verhoogd risico vormen voor het ontwikkelen van de ZVC, mede door het dragen van *CARD15* mutaties.

We stelden ons de vraag of de darm bij SpA patiënten verschillen vertoont in genexpressie in vergelijking met controles. In Chapter 7 beschrijven we afwijkingen in genexpressie van normale colon biopten bij SpA patiënten. Deze afwijkingen zijn bij patiënten met chronische darminflammatie te vergelijken met afwijkingen in de ZVC, wat de originele klinische data bevestigen. SpA kan daarom dienen als een pre-crohn model, ook bij het selecteren van kandidaatgenen voor darminflammatie.
Curriculum Vitae

Debby Laukens (°13/08/1977, Antwerpen, Belgium)

AFFILIATION

Department for Molecular Biomedical Research
Ghent University
Technologiepark 927, B-9052 Zwijnaarde
Tel: +329/ 331 36 64
Fax: +329/ 331 3609
E-mail: debby.laukens@dmbr.ugent.be

PRESENT AFFILIATION

Department of Gastroenterology
Ghent University Hospital
De Pintelaan 185, B-9000 Gent
Tel: +329/ 240 20 64
Fax: +329/ 240 49 84
E-mail: debby.laukens@ugent.be

EDUCATION AND ACADEMIC CAREER

1995-1997: Candidate (Bachelor) in Biology (University of Antwerp, RUCA)
1998-1999: Licentiate (Master) in Biochemistry (University of Antwerp, UA)

Master thesis: “Localization and characterization of ataxin-7 in the mouse” at the laboratory of Molecular Genetics, University of Antwerp.
Promoters: Claire Mauger and Christine Van Broeckhoven

1999-2000: PhD fellowship at the laboratory of Molecular Genetics, University of Antwerp and Flanders Interuniversity Institute for Biotechnology
Promoters: Claire Mauger and Christine Van Broeckhoven

2000-present: PhD fellowship at the Department for Molecular Biomedical Research, University of Ghent and the Flanders Interuniversity Institute for Biotechnology.
Promoters: Pieter Rottiers and Erik Remaut
Curriculum Vitae

SCIENTIFIC ACTIVITIES

Publications in international scientific journals


Abstracts in international scientific journals

**Participation of international meetings**

Symposium on auto-immunity, Antwerpen, Belgium, November 30, 2001  
*Participant*

3rd International congress on spondyloarthropathies, Gent, Belgium, October 2-5, 2002  
*Participant*

The regulation of mucosal inflammation, Keystone, Colorado, USA, April 1-6, 2003  
*Oral presentation* “CARD15 mutations in patients with spondyloarthropathy are linked with disease progression and evolution to Crohn’s disease.”

*Participant*

12th International congress of immunology, Montreal, Canada, July 19-22, 2004  
*Oral presentation* “Transcriptome analysis of pre-crohn patients in a cohort of spondyloarthropathy patients.”

4th International congress on spondyloarthropathy, Gent, Belgium, October 8, 2004  
*Poster presentation* “Transcriptome analysis of pre-crohn patients in a cohort of spondyloarthropathy patients.”

Inflammatory bowel disease, research drives clinics, Munster, Germany, September 2-3, 2005  
*Poster presentation* “Identification of susceptibility genes for Crohn's disease via transcriptome analysis.”

Digestive disease week 2006, Los Angeles, California, USA, May 20-25, 2006  
*Poster presentation* “Identification of susceptibility genes for Crohn's disease via transcriptome analysis.”

**Participation of national meetings**

3rd VIB seminar, Mol, Belgium, February 17-18, 2000

4th VIB seminar, Blankenberge, Belgium, February, March 15-16, 2001

5th VIB seminar, Blankenberge, Belgium, March 08-09, 2002

6th VIB seminar, Blankenberge, Belgium, March 13-14, 2003  
*Poster presentation* “CARD15 mutations in patients with spondyloarthropathy are linked with disease progression and evolution to Crohn’s disease.”
Doctoraatssymposium, ICC, Gent, Belgium, April 30, 2003
Poster presentation “CARD15 mutations in patients with spondyloarthropathy are linked with disease progression and evolution to Crohn’s disease.”

7th VIB seminar, Blankenberge, Belgium, March 11-12, 2004
Participant

4th VIB MicroArray user group meeting, Brussel, Belgium, November 19, 2004
Oral presentation “Identification of susceptibility genes for Crohn's disease via transcriptome analysis.”

8th VIB seminar, Blankenberge, Belgium, March 3-4, 2005
Participant

9th VIB seminar, Blankenberge, Belgium, March 09-10, 2006
Oral presentation “Alternative method to identify susceptibility genes for Crohn’s disease through transcriptome analysis.”

Awards

NIH grant award (Department of Health and Human Services, National Institutes of Health, National Institute of Allergy and Infectious Disease: Grant Number 1R13 AI 053891-01) for attending the meeting entitled “The Regulation of Mucosal Inflammation”, Keystone Resort, Colorado, USA.

FOCIS 2004 Travel Award (Federation of Clinical Immunology Societies), abstract submission entitled “Transcriptome Analysis of Pre-Crohn Patients in a Cohort of Spondyloarthropathy Patients.” 12th International congress for Immunology, Montreal, Canada.
De zoveelste rit van Gent naar huis, onderweg is er dan altijd veel tijd om na te denken. Ik voel een emotionele bui aankomen - er staat dan ook vanalles te veranderen - en ik besluit om mijn dankwoord te schrijven van zodra ik thuiskom. Zonet had ik nog een fijn gesprek met Sarah, dat gebeurt wel vaker. Zij vertrekt binnenkort naar Madrid, ik wens haar het allerbeste en zal ze missen! Ik ben thuisgekomen. Lichten aan, verwarming aan, de poes die rond mijn benen draait en mij zoals gewoonlijk bijna doet struikelen. Chris, je bent er niet vanavond. Ik heb honger, het is al 20u, en ik heb geen zin om te koken. Een beefje soep opwarmen en een pizza in de oven. Ik zit ongeveer halverwege mijn doctoraat, en ik vraag mij af waarom ik het toch zo moeilijk heb om door te werken, en alles op tijd af te krijgen. Gisteren kreeg ik te horen van Martine De Vos dat ik moet leren om meer gestructureerd te zijn, en ze heeft overschot van gelijk. Binnenkort kan ik beginnen in haar groep, ik kijk er naar uit, vooral ook omdat ik dan zal gaan samenwerken met Sara, die ondertussen een goede vriendin is geworden. Het klikt wel tussen ons, dus ik heb volledig het vertrouwen in onze samenwerking. Ook al zijn we allebei leeuwtjes - het zal er dus af en toe tegen zitten - we gaan er samen iets van maken! Voeg daarbij nog iemand zoals Harald, de wervelwind, en het is af. Harald, ook al ben je een geweldige arts, verlies ons en de research niet uit het oog! Bedankt voor het verzamelen van de honderden stalen en gegevens...

Vandaag heeft Hilde me er meermaals op gewezen dat ik toch eens zal moeten verder schrijven. Hilde ga ik zeker missen, ze is een geweldige steun en luisterend oor. Ik bedenk zonet dat ik misschien niet erg doordoe, omdat dit het afsluiten betekent van een periode waarvan ik echt heb genoten. Er is bijna geen dag geweest dat ik niet graag ging werken. Ik herinner mij Montreal. Samen met Ken, Jeroen en Pieter zitten we op een terras, en we drinken een pintje. Wat een fijn congres was dat! Ik kan niet goed tegen verandering, dat is altijd zo geweest. Nostalgie...

Er is zoveel gebeurd, zoveel gezegd, en ik heb zoveel geleerd. Ik heb dan ook veel nood aan discussie, net zoveel als dat ik nood heb om aan wetenschap te doen. De discussie is belangrijk, en ik vrees dat ik er meerderen mee heb verveeld. Vandaar dat dit tegelijk een “sorry, maar toch merci” moet worden.

Pieter, als men ooit een een positivo-gen gaat cloneren, zal jij de power zeker en vast verhogen. Bedankt voor de begeleiding, de steun en de immer positieve sfeer. Trouwens, ik wens jou, maar natuurlijk ook Lothar en Sabine, veel succes in het nieuwe bedrijf, en verzorg ons Hilde en Inge daar als prinseskes. Inge, de reddende engel en paniek-demper, altijd in de buurt om te helpen als er iets misgaat, en er is wat misgegaan, he! Weet je nog dat ik mijn cellen op de grond liet vallen? Gewoon opzuigen met een pasteurpipet en verder doen, zei ze. Het was nog goed afgelopen ook...
**Dankwoord**

**Erik Remaut** zal ik nooit vergeten. Iemand (ik weet natuurlijk niet meer wie) zei ooit: “*Active science narrows the mind more often than it broadens it*”, iets dat ik probeer in gedachte te houden. Erik is het levende bewijs dat dit niet altijd hoeft te gelden. Ik heb een mateloos respect voor wetenschappers die niet door allerlei systemen zijn aangetast, en die buiten een groot verstand, ook een groot hart hebben. Erik is ook de man die in 1 zin (al is die soms lang) kritisch kan samenvatten, en het voorbije uur vergaderen overbodig had kunnen maken. Erik, ik wens je al het allerbeste, geniet van je pensioen.

En dan ons **Michaela**, dedju toch, ons Michke. Waarom toch, he? Hoe komt het dat als er één persoon is op de wereld die absoluut een doctoraat moet verdienen, juist die er het hardst voor moet knokken. Je bent dan ook nog eens een mooi mens, verlies de moed niet, het tij zal wel keren... Samen met ons Karolien moeten je nog unit 1 recht houden, he! Veel plezier, en als je nog eens aan mij denkt, ik ben altijd bereid om gezellig samen te gaan eten... **Karolien**, onze gesprekjes over de zo vaak gelijkaardige “ingebeelde” kwaaltjes en angsten, en natuurlijk over de poes zal ik missen. Nog veel succes met je doctoraat, hopelijk gauw jou feestje? Ik denk nu ook aan **Veerle**. Veel geluk met je nieuwe kleine “projectje”, en natuurlijk ook met al de rest. Ons Ve, altijd goedlachs en fijn om mee te kletsen. Dat geldt natuurlijk ook voor **An**, altijd *in the mood* om eens goed te lachen, en altijd bereid om te helpen met een probleem. **Klaas** en **Sofie**, nog veel geluk samen, en, Sofie, ik hoop dat je snel een fijne job vindt! Ik vergeet nog bijna onze **Koen**! 'T Is ook zo’n stilleke, alhoewel... Stille waters? Veel succes op je nieuwe job, en ik hoop dat er gauw een tof vrouwke passeert die je kan thuis houden!

Dan zijn er nog de twee Tommen. **Tombo**, ik hoop dat je de moed vindt om eens eindelijk al je data samen te bundelen in een doctoraat, want dat is nu het enige dat telt voor jou. Je hebt zoveel in je mars. Ik heb veel aan jou te danken, zeker in het begin van mijn doctoraat. Denk nu eerst aan jezelf. **Tom** is ondertussen vertrokken naar het buitenland. Ik hoop dat alles daar verloopt zoals gepland en dat hij daar met zijn vrouwje een fantastische tijd beleeft. Tom, bedankt voor de leuke tijd! **Steven**, ik ben je inderdaad zeer dankbaar dat je mijn data hebt gered op een kritiek moment. **Miekje**, bedankt voor al die sequenties en hulp. Verder mag ik zeker niet vergeten om de **GOA** te bedanken. Deze maandelijkse vergaderingen waren erg belangrijk in het verloop van mijn project. Ook de **microarray faciliteit**, **applied-maths** en de **genetica faciliteit** hebben heel wat van het werk lichter gemaakt. Ondertussen beginnen we met *the girls* van de **pathologie** wat nauwer samen te werken. Ik denk dat we elkaar nog veel gaan nodig hebben, en **Anouk**: de metallothioneines pakken we samen aan he!

De laatste jaren heb ik vaak de opmerking gekregen dat ik wat meer moet buiten komen. ** Liesbetje**, bedankt voor het nalezen van mijn tekstjes, en ik hoop dat we binnenkort elkaar terug

172
wat meer zien! Dat geldt ook voor mijn lieve vriendin Babske. Die sauna-dagen moeten we meer doen, of... onze mannen mogen meer verrassingen organiseren!

Chrisje, ik heb zo de indruk dat je hebt geleerd om te gaan met mijn verschrikkelijke gemoedsschommelingen ten tijde van stress. Bedankt daarvoor, en voor de steun en knuffels!

Terwijl ik mij gedurende de laatste vijf jaar heb afvraagd wanneer ik aan kinderen moet beginnen, heeft mijn lieve broer en zijn Anneke er twee op de wereld gezet. Eerst “miniatuur Dirk” Milan en twee jaar later “nieuwe man” Patje. Wat een plezier die twee brengen is geweldig. Mercikes, Dirk en Anneke, dat jullie er zijn.

Mijn lieve bomma, die zo trots is op mij, moet ik ook beloven om eens wat meer langs te komen. Ik ga daar zeker werk van maken!

En dan, heb... papa en mama... Naast het feit dat jullie een fantastische dochter op de wereld hebben gezet ☺, is er zoveel om jullie voor te bedanken. Ik denk nog altijd dat ik de beste ouders heb die er ooit hebben bestaan. Papa, de cliche dat “mijne papa alles kan” is zo waar. Ik kan altijd bij jou terecht, ik kijk echt op naar jou. Mama, ik weet dat je erg fier bent op mij. Bedankt voor de steun, liefde en inspiratie...

Juni 2006