GENETIC VARIATION AND CLINICAL MANIFESTATIONS IN INFLAMMATORY BOWEL DISEASE

Maarit Lappalainen

Department of Medicine and
Research Program for Molecular Medicine,
Biomedicum Helsinki,
University of Helsinki, Finland

ACADEMIC DISSERTATION

To be publicly discussed with the permission of the Faculty of Medicine, University of Helsinki, in Lecture Hall 2, Haartman Institute, Haartmaninkatu 3, on 23rd May 2008, at 12 o’clock.

Helsinki 2008
# TABLE OF CONTENTS

## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF ORIGINAL PUBLICATIONS</td>
<td>5</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>6</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>7</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>9</td>
</tr>
<tr>
<td>REVIEW OF THE LITERATURE</td>
<td>10</td>
</tr>
<tr>
<td>1. INFLAMMATORY BOWEL DISEASE</td>
<td>10</td>
</tr>
<tr>
<td>1.1 Crohn’s disease</td>
<td>11</td>
</tr>
<tr>
<td>1.2 Ulcerative colitis</td>
<td>12</td>
</tr>
<tr>
<td>1.3 Diagnosis</td>
<td>13</td>
</tr>
<tr>
<td>1.4 Treatment</td>
<td>14</td>
</tr>
<tr>
<td>1.5 Epidemiology</td>
<td>15</td>
</tr>
<tr>
<td>2. GENETIC EPIDEMIOLOGY</td>
<td>16</td>
</tr>
<tr>
<td>2.1 Twin studies</td>
<td>16</td>
</tr>
<tr>
<td>2.2 Family studies</td>
<td>17</td>
</tr>
<tr>
<td>2.3 Racial and ethnic groups</td>
<td>18</td>
</tr>
<tr>
<td>2.4 Interaction between genes and environment</td>
<td>18</td>
</tr>
<tr>
<td>3. MOLECULAR GENETIC STUDIES</td>
<td>19</td>
</tr>
<tr>
<td>3.1 Candidate gene approach</td>
<td>19</td>
</tr>
<tr>
<td>3.2 Genome-wide linkage studies</td>
<td>20</td>
</tr>
<tr>
<td>3.2.1 Genome-wide linkage study in Finnish IBD families</td>
<td>21</td>
</tr>
<tr>
<td>3.3 Genome-wide association studies</td>
<td>22</td>
</tr>
<tr>
<td>4. IBD SUSCEPTIBILITY GENES</td>
<td>23</td>
</tr>
<tr>
<td>4.1 CARD15: the first CD susceptibility gene</td>
<td>26</td>
</tr>
<tr>
<td>4.1.1 CARD15 structure, expression, and function</td>
<td>26</td>
</tr>
<tr>
<td>4.1.2 NOD2 signalling pathway</td>
<td>27</td>
</tr>
<tr>
<td>4.1.3 CARD15 mutation frequency and genotype-phenotype associations in different CD cohorts</td>
<td>29</td>
</tr>
<tr>
<td>4.1.4 Rare CARD15 variants and their effect on phenotype</td>
<td>31</td>
</tr>
<tr>
<td>4.1.5 Functional consequences of CARD15 mutations</td>
<td>35</td>
</tr>
<tr>
<td>4.1.6 CARD15 in Blau syndrome and early-onset sarcoidosis</td>
<td>37</td>
</tr>
<tr>
<td>4.2 The DLD5 gene on chromosome 10</td>
<td>38</td>
</tr>
<tr>
<td>4.3 The ABCB1/MDR1 gene on chromosome 7</td>
<td>39</td>
</tr>
<tr>
<td>4.4 The TLR4 gene on chromosome 9</td>
<td>40</td>
</tr>
<tr>
<td>4.5 The TRAF6 gene on chromosome 11</td>
<td>41</td>
</tr>
<tr>
<td>4.6 The TNFRSF1A gene on chromosome 12</td>
<td>41</td>
</tr>
<tr>
<td>4.7 The IL23R gene on chromosome 1</td>
<td>42</td>
</tr>
<tr>
<td>4.8 The ATG16L1 gene on chromosome 2</td>
<td>43</td>
</tr>
<tr>
<td>5. STUDIES OF SPECIFIC GENE LOCI</td>
<td>44</td>
</tr>
<tr>
<td>5.1 Chromosome 5 risk haplotype (IBD5 locus)</td>
<td>44</td>
</tr>
<tr>
<td>5.2 Major histocompatibility complex region (IBD3 locus)</td>
<td>45</td>
</tr>
<tr>
<td>5.2.2 The HLA-DRB1 gene association with IBD</td>
<td>46</td>
</tr>
<tr>
<td>AIMS OF THE STUDY</td>
<td>48</td>
</tr>
<tr>
<td>PATIENTS AND METHODS</td>
<td>49</td>
</tr>
<tr>
<td>1. STUDY SUBJECTS</td>
<td>49</td>
</tr>
<tr>
<td>1.1 Patients with inflammatory bowel disease</td>
<td>49</td>
</tr>
<tr>
<td>1.2 Healthy controls</td>
<td>50</td>
</tr>
<tr>
<td>1.3 Ethical aspects</td>
<td>50</td>
</tr>
<tr>
<td>2. METHODS</td>
<td>51</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS

RESULTS...................................................................................................................................................... 52
1. MUTATIONS IN THE CARD15 GENE AND THEIR RELATION TO CLINICAL PHENOTYPES. 52
   1.1 The CARD15 gene mutation spectrum (I, II) ................................................................................... 52
   1.2 The CARD15 mutation frequency in the background population (III) ........................................ 55
   1.3 Phenotypic characteristics of CARD15 mutations (I, II) ................................................................ 56
2. FUNCTIONAL STUDIES OF THE CARD15 GENE (II) ..................................................................... 57
3. GENETIC VARIATIONS IN CANDIDATE GENES AND CORRELATION TO PHENOTYPE...... 60
   3.1 Screening of the TRAF6 gene (IV) .................................................................................................... 60
   3.2 The rare HLA-DRB1*0103 allele (V) .............................................................................................. 60
   3.3 Eight SNPs spanning the IL23R gene (V) ........................................................................................... 60
   3.4 Three SNPs within the TNFRSF1A gene (V) ................................................................................... 61
   3.5 SNPs in genes ATG16L1, DLG5, ABCB1, TLR4, and the chromosome 5 risk haplotype containing SLC22A4 and SLC22A5 (V) ................................................................................... 62

DISCUSSION................................................................................................................................................ 63
1. GENETIC VARIATIONS IN THE CARD15 GENE ............................................................................. 63
   1.1 CARD15 mutation frequency in the background population ............................................................ 66
   1.2 Genotype-phenotype correlations of CARD15 mutations ............................................................... 67
2. GENETIC VARIATION IN OTHER IBD CANDIDATE GENES....................................................... 68
   2.1 Positive association with IL23R, TNFRSF1A, and HLA-DRB1*0103 allele variants ................. 68
   2.2 No association with TRAF6, ATG16L1, DLG5, ABCB1, TLR4, SLC22A4, SLC22A5, or the chromosome 5 haplotype ........................................................................................................................ 69

CONCLUDING REMARKS AND FUTURE PROSPECTS............................................................................ 71

ACKNOWLEDGEMENTS ............................................................................................................................ 72

REFERENCES ............................................................................................................................................. 74
LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications:


The publications are referred to in the text by their Roman numerals I-V and are reproduced with the permission of the copyright holders.
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCB1</td>
<td>ATP-binding cassette, subfamily B, member 1</td>
</tr>
<tr>
<td>ASCA</td>
<td>anti-<em>Saccharomyces cerevisiae</em> antibody</td>
</tr>
<tr>
<td>ATG16L1</td>
<td>autophagy 16-like 1</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>CARD15</td>
<td>caspase recruitment domain-containing protein 15</td>
</tr>
<tr>
<td>CD</td>
<td>Crohn’s disease</td>
</tr>
<tr>
<td>DLG5</td>
<td>discs large, Drosophila, homolog of, 5</td>
</tr>
<tr>
<td>CRP</td>
<td>C-reactive protein</td>
</tr>
<tr>
<td>ELISA</td>
<td>enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>ESR</td>
<td>erythrocyte sedimentation rate</td>
</tr>
<tr>
<td>GI</td>
<td>gastrointestinal</td>
</tr>
<tr>
<td>GWA</td>
<td>genome-wide association</td>
</tr>
<tr>
<td>dHPLC</td>
<td>denaturing high-performance liquid chromatography</td>
</tr>
<tr>
<td>HEK</td>
<td>human embryonic kidney</td>
</tr>
<tr>
<td>HLA</td>
<td>human leukocyte antigen</td>
</tr>
<tr>
<td>IBD</td>
<td>inflammatory bowel disease</td>
</tr>
<tr>
<td>IL23R</td>
<td>interleukin 23 receptor</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>LD</td>
<td>linkage disequilibrium</td>
</tr>
<tr>
<td>LRR</td>
<td>leucine-rich repeat</td>
</tr>
<tr>
<td>Mb</td>
<td>megabase</td>
</tr>
<tr>
<td>MDP</td>
<td>muramyl dipeptide</td>
</tr>
<tr>
<td>MDR1</td>
<td>multidrug resistance 1</td>
</tr>
<tr>
<td>MHC</td>
<td>major histocompatibility complex</td>
</tr>
<tr>
<td>NBD</td>
<td>nucleotide binding domain</td>
</tr>
<tr>
<td>NOD2</td>
<td>nucleotide-binding oligomerization domain protein 2</td>
</tr>
<tr>
<td>NLR</td>
<td>NOD-like receptor</td>
</tr>
<tr>
<td>NPL</td>
<td>nonparametric linkage</td>
</tr>
<tr>
<td>OCTN1</td>
<td>organic cation transporter 1</td>
</tr>
<tr>
<td>OCTN2</td>
<td>organic cation transporter 2</td>
</tr>
<tr>
<td>pANCA</td>
<td>perinuclear anti-neutrophil cytoplasmic antibody</td>
</tr>
<tr>
<td>PAMP</td>
<td>pathogen-associated molecular pattern</td>
</tr>
<tr>
<td>PGN</td>
<td>peptidoglycan</td>
</tr>
<tr>
<td>PIRA</td>
<td>primer-induced restriction assay</td>
</tr>
<tr>
<td>PRR</td>
<td>pattern recognition receptor</td>
</tr>
<tr>
<td>RFLP</td>
<td>restriction fragment length polymorphism</td>
</tr>
<tr>
<td>SLC22A4</td>
<td>solute carrier family 22, member 4</td>
</tr>
<tr>
<td>SLC22A5</td>
<td>solute carrier family 22, member 5</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>TLR4</td>
<td>toll-like receptor 4</td>
</tr>
<tr>
<td>TNFRSF1A</td>
<td>tumor necrosis factor receptor superfamily, member 1a</td>
</tr>
<tr>
<td>TNFSF15</td>
<td>tumor necrosis factor superfamily, member 15</td>
</tr>
<tr>
<td>TRAF6</td>
<td>TNF receptor-associated factor 6</td>
</tr>
<tr>
<td>UC</td>
<td>ulcerative colitis</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
</tbody>
</table>

In addition, standard one-letter abbreviations are used for nucleotides and amino acids.
ABSTRACT

Crohn’s disease (CD) and ulcerative colitis (UC), two major forms of inflammatory bowel disease (IBD), are chronic relapsing conditions of the gastrointestinal tract. IBD prevalence in Finland is approximately 3-4 per 1000 and the peak age of onset is between 20 and 35 years of age. The precise aetiology of IBD is unknown but interplay of environmental risk factors and immunologic changes trigger the disease in a genetically susceptible individual. Twin and family studies have provided strong evidence for genetic factors in IBD susceptibility, and genetic factors may be more prominent in CD than UC. The first CD susceptibility gene was identified in 2001. Three common mutations R702W, G908R, and 1007fs of the \textit{CARD15/NOD2} gene are shown to associate independently with CD but the magnitude of association varies between different populations.

One of the main objectives of this study was to evaluate the role of \textit{CARD15} in a Finnish CD cohort. 271 CD patients were studied for the three common mutations and the results showed a lower mutation frequency than in other Caucasian populations. Only 16% of the patients carried one of the three mutations. Ileal location as well as stricturing and penetrating behaviour of the disease were associated with occurrence of the mutations. The whole protein coding region of \textit{CARD15} was screened for possible Finnish founder mutations. In addition to several sequence variants, five novel mutations (R38M, W355X, P727L, W907R, and R1019X) were identified in five patients. Functional consequences of these novel variants were studied \textit{in vitro}, and these studies demonstrated a profound impairment of MDP response. Investigation of \textit{CARD15} mutation frequency in healthy people across three continents showed a large geographic fluctuation. No simple correlation between mutation frequency and disease incidence was seen in populations studied. The occurrence of double mutant carriers in healthy controls suggested that the penetrance of risk alleles is low.

Other main objectives aimed at identifying other genetic variations that are involved in the susceptibility to IBD. We investigated the most plausible IBD candidate genes including \textit{TRAF6}, \textit{SLC22A4}, \textit{SLC22A5}, \textit{DLG5}, \textit{TLR4}, \textit{TNFRSF1A}, \textit{ABCB1/MDR1}, \textit{IL23R}, and \textit{ATG16L1}. The marker for a chromosome 5 risk haplotype and the rare HLA-DRB1*0103 allele were also studied. The study cohort consisted of 699 IBD patients (240 CD and 459 UC), of which 23% had a first-degree relative with IBD. Of the several candidate genes studied, \textit{IL23R} was associated with CD susceptibility, and \textit{TNFRSF1A} as well as the HLA-DRB1*0103 allele with UC susceptibility. \textit{IL23R} variants also showed association with the stricturing phenotype and longer disease duration in CD patients. In addition, \textit{TNFRSF1A} variants were more common among familial UC and ileocolonic CD.
In conclusion, the common *CARD15* mutations were shown to account for 16% of CD cases in Finland. Novel *CARD15* variants identified in the present study are most likely disease-causing mutations, as judged by the results of *in vitro* studies. The present study also confirms the *IL23R* association with CD susceptibility and, in addition, *TNFRSF1A* and HLA-DRB1*0103 allele association with UC of specific clinical phenotypes.
INTRODUCTION

Inflammatory bowel disease or IBD usually refers to either Crohn’s disease (CD) or ulcerative colitis (UC). Case descriptions of diseases that today would probably be called IBD were published as far back as several centuries ago (Russel 2000). In 1913, doctor Thomas Kennedy Dalziel reported thirteen patients suffering from intestinal obstruction in the *British Medical Journal* (Hawkins 1990). On autopsy he found that all patients had an inflamed gut and the inflammation was transmural, which is today known as one of the characteristics of CD. Between the 1920s and 1930s more young people with CD like symptoms were seen. In 1932, doctors Crohn, Ginsberg, and Oppenheimer published their findings in *The Journal of the American Medical Association*, and mainly because of the timing CD came to be known as a medical entity.

In 1859 Sir Samuel Wilks described the first case of UC in a letter to the *Medical Times and Gazette* even though at that time he did not know the correct diagnosis (Hawkins 1990). Later, in 1875, Wilks with Walter Moxon described UC showing that the disease had a non-infectious cause. Already at the beginning it was thought that CD and UC are two distinct conditions. At present, in about 10% of cases the distinction between these two disorders can not be made and the term indeterminate colitis is used.

Before the human genome was sequenced mainly the genes responsible for rare Mendelian disorders were successfully found. It has been more difficult to find disease-causing gene variants for complex diseases like IBD because the effect of individual genes is small and gene interactions play an essential role. In addition, environmental factors are important in disease development. In genetics, recent advances in genotyping technologies and determination of the human genome sequence have made a huge impact on resolving the genetic causes of common disorders. Establishment of repositories of large patient samples is also a powerful tool in common disease genetics. Twin and family studies have shown that genetic susceptibility is essential in IBD development, nevertheless only a few IBD susceptibility genes have been identified to date.

In the present study, genetics of two inflammatory bowel diseases, Crohn’s disease and ulcerative colitis, and possible correlation of genetic variations to clinical characteristics were investigated in an IBD cohort of Finnish origin.
REVIEW OF THE LITERATURE

1. INFLAMMATORY BOWEL DISEASE

Inflammatory bowel disease (IBD) is a group of inflammatory conditions affecting the gastrointestinal (GI) tract. The two main forms of IBD are Crohn’s disease (CD) and ulcerative colitis (UC). Typically, the onset of IBD is in late adolescence or early adulthood (Russel and Stockbrugger 1996; Farrokhyar et al. 2001) with equal ratios of men and women in most reports. The symptoms of IBD include diarrhoea, abdominal pain, fever, and weight loss. However, CD and UC can usually be distinguished by clinical and pathologic characteristics. The main difference between these two disorders is the location and nature of the inflammation (Table 1) (Bouma and Strober 2003). In CD, the inflammation can involve any part of the GI tract from mouth to anus, but it most commonly affects the small intestine (jejunum and/or ileum) and colon. UC instead affects only the colonic part of the GI tract, sometimes only the rectum. Microscopically, inflammation in CD is transmural affecting the whole bowel wall, while in UC it is restricted to the mucosal layer.

Table 1. Differentiation between Crohn’s disease and ulcerative colitis (Lennard-Jones 1989; Baumgart and Sandborn 2007).

<table>
<thead>
<tr>
<th>Clinical features</th>
<th>Crohn’s disease</th>
<th>Ulcerative colitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon involvement</td>
<td>Usually</td>
<td>Always</td>
</tr>
<tr>
<td>Terminal ileum involvement</td>
<td>Commonly</td>
<td>Seldom</td>
</tr>
<tr>
<td>Rectum involvement</td>
<td>Seldom</td>
<td>Usually</td>
</tr>
<tr>
<td>Involvement around anus</td>
<td>Common</td>
<td>Seldom</td>
</tr>
<tr>
<td>Distribution of inflammation</td>
<td>Patchy</td>
<td>Continuous</td>
</tr>
<tr>
<td>Nature of inflammation</td>
<td>Transmural</td>
<td>Superficial</td>
</tr>
<tr>
<td>Ulcers in endoscopy</td>
<td>Serpiginous and Deep</td>
<td>Superficial</td>
</tr>
<tr>
<td>Fistulas</td>
<td>Common</td>
<td>Seldom</td>
</tr>
<tr>
<td>Stenosis</td>
<td>Common</td>
<td>Seldom</td>
</tr>
<tr>
<td>Crypt abscesses on biopsy</td>
<td>Uncommon</td>
<td>Common/Always</td>
</tr>
<tr>
<td>Granulomas on biopsy</td>
<td>Common</td>
<td>Uncommon</td>
</tr>
</tbody>
</table>

Both diseases may have similar extraintestinal manifestations usually affecting liver (primary sclerosing cholangitis, autoimmune hepatitis), skin (erythema nodosum, pyoderma gangrenosum), eyes (uveitis, iritis), and joints (arthritis, ankylosing spondylitis) (Rothfuss et al. 2006). In about 10% of IBD cases it is not possible to distinguish between CD and UC, and then the term indeterminate colitis is used.
1.1 Crohn’s disease

Crohn’s disease (CD, also known as regional enteritis, granulomatous ileitis, and ileocolitis) is named after American gastroenterologist Burrill B. Crohn. In CD, the main symptoms include abdominal pain, diarrhoea, bloody stools, pain in defecation, rectal bleeding and pain, fever, and weight loss (Farmer et al. 1975; Hanauer and Sandborn 2001; Carter et al. 2004). Many patients have symptoms for years before a diagnosis is made. The symptoms also vary between patients depending on the part of the gut affected. Phenotypic classification of the disease is based on the location and behaviour of the inflammation. Vienna classification is a simple classification of CD based on a set of objective variables including age of onset (Table 2) (Gasche et al. 2000). In almost half of the cases, CD affects both the ileum and the colon (ileocolitis) (Farmer et al. 1985). In around thirty percent of cases only the small bowel (jejunileitis or ileitis) is affected and in the remaining twenty percent the large intestine (colitis) is affected. The upper GI tract, esophagus, stomach, or duodenum is affected in about ten percent of cases with or without the inflammation in the small intestine or colon. Behaviour of this condition can be divided into three categories: stricturing, penetrating, and inflammatory. Stricturing behaviour (stenosis) causes narrowing of the bowel that may lead to bowel obstruction. The penetrating form of the disease creates fistulas between the bowel and other organs such as the skin. Fistulas occur most commonly around the anal area. Patients with a pure inflammatory form of CD do not develop either fistulas or strictures. The anatomical location of CD is fairly stable whereas the behaviour varies substantially during the natural disease course (Baumgart and Sandborn 2007).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
</table>
| Age at onset | 1. < 40 years  
2. ≥ 40 years |
| Disease location | 1. terminal ileum  
2. colon  
3. ileocolon  
4. upper gastrointestinal tract |
| Disease behaviour | 1. inflammatory  
2. stricturing  
3. penetrating |

CD activity indexes have been developed to standardise and quantify disease severity. Crohn’s Disease Activity Index (CDAI) from 1976 uses eight different clinical variables in the determination of disease activity (Best et al. 1976). Index values below 150 indicate remission and values above 450 are an indication of severe disease. A few years later the simplified CDAI was developed (Harvey and Bradshaw 1980). This simplified index,
Harvey–Bradshaw Index (HBI, Simple Index) is based on only five variables. Both the Vienna classification and the CDAI/HBI are mainly used in scientific research for the determination of disease severity. Most recently, fecal calprotectin and lactoferrin have been shown to be more sensitive surrogate markers than normally used activity indexes (Sipponen et al. 2008). Regarding disease activity, only 10-13% of patients will remain in remission for several years (Munkholm et al. 1995; Loftus et al. 2002). About 70% of patients have a chronic intermittent course while in the remainder the disease is constantly active.

1.2 Ulcerative colitis

In ulcerative colitis (UC) the colon is affected by the formation of ulcers. The patients usually present with diarrhoea mixed with blood and mucus and sometimes also abdominal pain, urgency, or tenesmus (Carter et al. 2004; Kornbluth and Sachar 2004; Baumgart and Sandborn 2007). As in CD, the symptoms may occur on and off for a long period of time. UC is also classified according to the extent of the inflammation. This depends on how far up the colon the disease extends. The inflammation usually begins from the rectum. In simple classification the disease is divided into a distal or extensive disease (Kornbluth and Sachar 2004). Distal colitis can be further divided into proctitis (only the rectum affected), proctosigmoiditis (the rectosigmoid colon affected), or left-sided colitis (the descending colon affected) (Baumgart and Sandborn 2007). In pancolitis, the inflammation covers the whole colon extending from the rectum to the cecum, and sometimes even further to the distal ileum – a phenomenon termed as backwash ileitis.

In addition to the extent of involvement, UC can be categorised by the severity of the disease. Disease severity is defined as mild, moderate, severe, or fulminant (Truelove and Witts 1955; Hanauer 1996; Carter et al. 2004; Silverberg et al. 2005). In mild cases, there are less than four stools per day, and in moderate disease stool frequency is more than four with minimal signs of toxicity. Severe UC is manifested by more than six bloody stools daily and toxicity is demonstrated by fever, tachycardia, anaemia, or an elevated erythrocyte sedimentation rate (ESR). Patients with fulminant disease usually have more than ten bowel movements per day with continuous bleeding, toxicity, abdominal tenderness and distension, and require blood transfusions and colonic dilation. During the natural course of the disease it may progress proximally (Langholz et al. 1994; Henriksen et al. 2006). Disease activity in the overall patient population is remarkably constant with about half of the patients in clinical remission at any given time, but the majority of patients have an intermittent course (Langholz et al. 1994). In addition, approximately
20% and 40% of UC patients did not have a relapse or were not taking medication, respectively, five years after diagnosis (Henriksen et al. 2006).

1.3 Diagnosis

There is no single test for IBD. Diagnosis is a combination of clinical symptoms, radiologic or endoscopic appearance of the bowel together with typical histology and exclusion of other causes of intestinal inflammation (Lennard-Jones 1989; Hanauer and Sandborn 2001; Baumgart and Sandborn 2007). Usually the diagnosis between CD and UC is clear but in approximately 10% of cases the distinction can not be made and the term indeterminate colitis is used. With time the symptoms or clinical findings become clearer which makes the differential diagnosis between these two conditions possible. Colonoscopy is a primary tool for diagnostics since it allows direct visualization from the rectum to the terminal ileum. A wireless capsule endoscopy can be used in CD diagnosis to evaluate the extent of the small bowel involvement. In this procedure, a small capsule with a built-in camera is swallowed and the camera takes serial pictures of the entire small bowel, finally ending up in the patient’s faeces. In CD, the inflammation is discontinuous whereas in UC it is continuous. In addition, deep ulcers may be present in CD.

During the endoscopy, several biopsies are taken from various parts of the bowel in order to confirm the diagnosis and to assess the inflammation activity. CD shows a transmural pattern of inflammation and granulomas are present in about 50% of CD cases. In UC histopathology, the inflammation is more diffuse and restricted to mucosal layers. The pathology typically involves distortion of crypt architecture and inflammation of crypts (cryptitis).

C-reactive protein (CRP) correlates well with deep inflammation in the bowel wall in CD (Vermeire et al. 2006). Bleeding and malabsorption cause anaemia, which may be due to iron, folate, or vitamin B_{12} deficiencies. Laboratory tests in mild and moderate UC cases are usually normal (Silverberg et al. 2005). Stool samples for calprotectin and lactoferrin measurements have been used in IBD diagnosis for a few years now (Sipponen et al. 2008). However, sometimes in moderate and always in the most severe cases blood tests show elevated acute-phase reactants CRP and ESR, leukocytosis, thrombocytosis, and hypoalbuminemia. Stool cultures need to be performed in both diseases to rule out parasites and infectious causes.

Although there are no specific diagnostic laboratory tests for IBD, UC and CD can be characterised by different serological antibody expression. High levels of perinuclear
antineutrophil cytoplasmic antibodies (pANCA) in the serum are strongly correlated with occurrence of UC (Shanahan et al. 1992; Seibold et al. 1994; Lee et al. 1995; Halme et al. 2002) while serum reactivity to anti-Saccharomyces cerevisiae antibodies (ASCA) is associated with CD (Lee et al. 1995; Quinton et al. 1998; Sutton et al. 2000; Seibold et al. 2001; Halme et al. 2002). However, these tests do not reliably separate the two diseases. They have uncertain value in indeterminate colitis cases and are not recommended for routine diagnosis.

1.4 Treatment

The treatment of IBD is sequential: first, the treatment of acute disease in order to achieve remission and then the maintenance of remission. In addition, minimising the side effects of the treatment and improving quality of life are other equally important factors. Remission is usually achieved by medication. Medical treatment is dependent on the disease course, extent of inflammation, and disease severity. Patients with IBD often require a combination of drugs to control their disease (Irving et al. 2008). However, in the mildest cases patients may not need any treatment. In both disorders, medical therapy consists of anti-inflammatory drugs such as 5-aminosalicylic acid (5-ASA) and corticosteroids as well as immunosuppressive agents (azathioprine or its derivatives, methotrexate, cyclosporine), sometimes also antibiotics. 5-ASA can be used in the treatment of mild IBD as well as in the induction and maintenance of remission (Carter et al. 2004; Kornbluth and Sachar 2004). Corticosteroids are effective in moderate to severe UC and in reducing the symptoms of CD. Because of side-effects, the prolonged use of corticosteroids should be avoided. Drug treatment in severe IBD consists of corticosteroids, which should be reduced over time and then later switch on to immunosuppressants, if needed. Although some patients are able to maintain remission with 5-ASA alone, many require immunosuppressive drugs (Hanauer and Sandborn 2001). The precise role of antibiotics in the treatment of IBD is unknown, however, in clinical practise, antibiotics are frequently used (Perencevich and Burakoff 2006). Antibiotics may be useful in the treatment of CD patients with colonic disease (Sutherland et al. 1991). The proven involvement of cytokines in the IBD pathophysiology has led to the development of anticytokine drugs, so-called biologics, as a therapy for IBD over the last decades. Among biological therapies, anti-TNF antibodies (infliximab or adalimumab) play a pivotal role in the treatment of chronic active IBD and fistulating disease when other medical treatments fail to induce remission (Kuhbacher and Folsch 2007).

Unresponsiveness to medical treatment and complications are the main reasons for surgery. Around 55-90% of CD patients ultimately require an operation (Farmer et al.
1976; Hanauer and Sandborn 2001; Larson and Pemberton 2004). Indications for surgery are usually intestinal obstruction, internal fistula and abscess, or poor response to medical therapy (Farmer et al. 1976). Perforation of the bowel and severe bleeding may sometimes necessitate surgery. Resections in CD should be, however, conservative (Carter et al. 2004). Resection of the involved bowel may improve the symptoms but does not cure the disease. Unlike CD, UC can be cured by surgically removing the large intestine and rectum (proctocolectomy). Most patients with UC will not require surgery but approximately one third will eventually need an operation (Larson and Pemberton 2004). Emergency colectomy is indicated in case of massive hemorrhage, severe toxic megacolon as a result of fulminanat colitis, or perforation (Berg et al. 2002; Cima and Pemberton 2005). Another indication for surgery is unresponsiveness to medication (Larson and Pemberton 2004; Cima and Pemberton 2005). In both disorders, documented or strongly suspected dysplasia and carcinoma are indications for elective surgery.

1.5 Epidemiology

The incidence of IBD has increased dramatically in the western world since the early 1950s (Russel 2000; Jess et al. 2007; Loftus et al. 2007), with a peak onset between 20 and 35 years of age. The highest incidence rates have been reported for North America, the United Kingdom, and Scandinavia, whereas lower rates are found in southern Europe, Africa, Asia, and South America (Farrokhyar et al. 2001; Loftus 2004; Lakatos 2006; Baumgart and Carding 2007). The overall annual incidence rate for CD varies between 0.1-16/100 000 inhabitants and for UC between 0.5-24.5/100 000 inhabitants (Ouyang et al. 2005; Jacobsen et al. 2006; Lakatos 2006; Lapidus 2006). The corresponding IBD prevalence reaches as high as 396/100 000 inhabitants (Ouyang et al. 2005; Lakatos 2006). CD and UC prevalence varies between 26-213/100 000 and 37-294/100 000 inhabitants, respectively (Loftus 2004; Jacobsen et al. 2006; Lapidus 2006). Geography has a strong influence on IBD incidence, with higher incidence in the northern than southern hemisphere (Garland et al. 1981; Sonnenberg et al. 1991; Sonnenberg and Wasserman 1991; Shivananda et al. 1996; Loftus et al. 1998; Bernstein et al. 1999; Loftus et al. 2000; Russel 2000). IBD is also more common in developed and industrialized countries as well as in urban areas (Russel and Stockbrugger 1996; Loftus et al. 2000; Russel 2000; Lakatos 2006; Baumgart and Carding 2007). It has recently been observed that the gap between areas of high and low incidences is diminishing, possibly as a result of a changing, more westernised, life-style (Russel 2000). It is also of note that CD incidence is still increasing in many areas whereas UC incidence has been stable for some time (Farrokhyar et al. 2001) and in some areas CD has become more prevalent than UC.
REVIEW OF THE LITERATURE

(Molinie et al. 2004). Despite the stabilisation of incidence, the prevalence still continue to grow as a result of the early age of onset and low mortality (Loftus et al. 2007).

In a European collaborative study during 1991-93, both UC and CD incidences were higher in the northern than southern Europe (Shivananda et al. 1996). According to this study the overall incidence in Europe was 10.4/100 000 individuals for UC and 5.6/100 000 for CD. UC rates in northern centres were 40% higher than those in the south and for CD they were up to 80% higher, but this gradient was narrower than previously believed. In addition, it seems that continuously increasing incidences are still reported in high incidence areas such as Denmark (Vind et al. 2006), Sweden (Lapidus 2006), and Finland (Turunen et al. 2006). In Finland, the IBD prevalence was 240/100 000 inhabitants in 1993 (Karvonen et al. 1993). In 2006, however, according to drug compensation statistics from the Social Insurance Institution of Finland (KELA), the IBD prevalence had increased to over 500/100 000. Among children, the IBD incidence has almost doubled in the last 20 years (Turunen et al. 2006), so it is obvious that the trend is similar among adults, although there are no reports on that. The annual CD incidence during 1996-2000 was 5.6/100 000 and the corresponding prevalence was 79.7/100 000 (Jussila et al. 2004). The latest UC incidence and prevalence rates are from 1993 when the corresponding rates were 8.1/100 000 and 147/100 000, respectively (Karvonen et al. 1993).

2. GENETIC EPIDEMIOLOGY

Epidemiological, clinical, and molecular studies have provided strong evidence that inherited predisposition is important in the pathogenesis of chronic IBD. IBD is a complex disorder with an aetiology that is only partly understood. In complex disorders, both genetic and environmental factors have a role in the development of the disease. Unlike Mendelian inheritance, in complex diseases genes are expected to have a low penetrance. Therefore, genes are less prone to selection which results in higher gene frequencies than expected. IBD is now widely believed to originate from a dysregulated immune response to luminal bacteria in a genetically susceptible individual (Abreu and Sparrow 2006; Bamias and Cominelli 2007; Van Limbergen et al. 2007b). The genetic predisposition in IBD has long been known since epidemiological studies have shown that the prevalence of IBD is increased among relatives with CD and UC.

2.1 Twin studies

Twin studies have been a powerful tool for the identification of different contributions of genes and environmental factors in the aetiology of complex diseases. If a disease was to
be entirely due to genes, like in monogenic diseases, the concordance would be 100% in identical (monozygotic) twins and 50% in non-identical (dizygotic) twins. In contrast, if the disease was entirely dependent on non-inherited factors, then the concordance would be similar in both twin pairs. Strong evidence for the involvement of genetic factors in IBD derives from twin studies which have shown significantly higher concordance rates in monozygotic than in dizygotic twins. Large twin studies from Sweden, Denmark, and the United Kingdom show that the concordance rate for CD was 20-58% in identical twins and less than 10% in non-identical twins when brought up in same environment (Tysk et al. 1988; Thompson et al. 1996; Orholm et al. 2000; Halfvarson et al. 2003; Russell and Satsangi 2004). The corresponding rates for UC were 6-19% in monozygotic twins and 4% in dizygotic twins (Halfvarson et al. 2003; Russell and Satsangi 2004). These numbers show that the genetic influence is stronger in CD than in UC.

2.2 Family studies

Since the first observation of familial occurrence in 1963, family studies have shown significant familial clustering of IBD (Kirsner and Spencer 1963; Farmer et al. 1980; Orholm et al. 1991; Satsangi et al. 1994; Peeters et al. 1996). A positive family history is the most important independent risk factor for the disease. In addition, the greatest risk is to develop the same type of IBD as the affected relative. Figures vary from one study to another, but approximately 10-15% of adult IBD patients have a family member affected with IBD (Bonen and Cho 2003). The risk is particularly strong among first-degree relatives. CD patients have a first-degree relative with CD in 2.2-16.2% of cases and with IBD in 5.2-22.5% of cases (Russell and Satsangi 2004). In UC patients, the corresponding figures are 5.7-15.5% with UC and 6.6-15.8% with IBD. At highest risk are children of two affected parents (>30%) (Laharie et al. 2001). A positive family history is more common in CD than in UC (Farmer et al. 1980; Satsangi et al. 1994). Earlier disease onset in the offspring of IBD patients is consistently found, and genetic anticipation has been hypothesized (Polito et al. 1996b). This phenomenon, however, may be a result of a combination of a time trend, increasing incidence of IBD. Mixed families also exist where both CD and UC occurs within the same family, at a frequency greater than expected by chance alone, indicating that these two are closely related disorders and may even share some susceptibility genes.

A degree of familial clustering of a disorder may be expressed as a ratio (λs) of a risk for siblings to the reported population prevalence. Using this λs value as a relative risk, the risk range is 15-42 for CD and 7-17 for UC (Satsangi et al. 1997; Ma et al. 1999; Ahmad et al. 2001). These λs values for CD are even higher than for many other complex
diseases, including type 1 and 2 diabetes (15 and <10), celiac disease (7-30), and schizophrenia (8.6) (Todd and Farrall 1996; Ahmad et al. 2001; Pena and Wijmenga 2001; Halme et al. 2006). This information shows that genes are a strong contributor to the development of CD in particular.

2.3 Racial and ethnic groups

Considerable differences exist in the incidence and prevalence of IBD between different racial and ethnic groups (Roth et al. 1989; Yang et al. 1993). The highest prevalence is seen among Caucasians, lower prevalence rates in black Americans, and the lowest rates in Hispanics and Asians (Satsangi et al. 1994; Duerr 2003; Baumgart and Carding 2007). Among Caucasians the highest incidence and prevalence rates are observed in Jewish populations with two to four times higher than in non-Jewish Whites (Yang et al. 1993; Polito et al. 1996a; Vermeire and Rutgeerts 2005). Within Jewish populations, rates of IBD are greater in Ashkenazi Jews than in Sephardic/Oriental Jews (Gilat et al. 1974; Rozen et al. 1979; Krawiec et al. 1984; Odes et al. 1987a; Odes et al. 1987b). And even among Ashkenazi Jews the rates for IBD are greater in Jews of central European origin than in those of Polish and Russian origin.

2.4 Interaction between genes and environment

IBD is currently believed to result from the complex effect of diverse genes conferring a risk of disease and environmental factors, which when combined lead to an aberrant inflammation response in the gut (Ahmad et al. 2001; Ferguson et al. 2007; Yamamoto-Furusho and Podolksy 2007). The discovery of the CARD15 gene in the CD pathogenesis (Hampe et al. 2001; Hugot et al. 2001; Ogura et al. 2001a) has refocused attention on innate immunity and the interaction between genetic factors and bacterial flora. The innate immune response is the first line of defence against invading pathogens (Aderem and Ulevitch 2000; Schwartz 2002; Bourhis et al. 2007) and the human intestinal tract is exposed to an enormous microbial flora all the time. The mucosal immune system has evolved to balance the need to respond to pathogens while co-existing with commensal bacteria and food antigens (Yamamoto-Furusho and Podolksy 2007). Innate immune responses are initiated by activation of pattern recognition receptors (PRRs), which consist of NOD-like receptor (NLR) and toll-like receptor (TLR) protein family members (Philpott and Girardin 2004; Bourhis et al. 2007; Kanneganti et al. 2007; Franchi et al. 2008). Immediate activation of innate immunity relies on detection of conserved microbial motifs known as pathogen associated molecular patterns (PAMPs), comprising diverse molecules from bacteria and viruses such as lipopolysaccharide (LPS), peptidoglycan
REVIEW OF THE LITERATURE

(PGN), flagellin, lipoprotein, and viral double-stranded RNA (Barton and Medzhitov 2003; Akira et al. 2006).

Of exogenous factors, smoking is the most consistent finding in IBD showing that smokers are at a higher risk of developing CD than non-smokers (Tobin et al. 1987; Lindberg et al. 1988; Logan and Kay 1989; Silverstein et al. 1989; Corrao et al. 1998; Halme et al. 2002; Garcia Rodriguez et al. 2005; Halfvarson et al. 2006). On the other hand, smoking has been shown to have a protective effect against UC (Benoni and Nilsson 1987; Tobin et al. 1987; Logan and Kay 1989; Rutgeerts et al. 1994; Parrello et al. 1997; Corrao et al. 1998; Halme et al. 2002; Garcia Rodriguez et al. 2005; Firouzi et al. 2006; Halfvarson et al. 2006; de Saussure et al. 2007). Some studies have also suggested appendectomy, use of oral contraceptives, breastfeeding, high sugar and fat diet, infections in childhood and vaccination as risk factors for IBD but the results are controversial (Farrokhyar et al. 2001; Loftus 2004).

3. MOLECULAR GENETIC STUDIES

3.1 Candidate gene approach

In the early years attempts to identify IBD susceptibility genes focused primarily on candidate gene studies. Today this approach is still used although more sophisticated methods have emerged. The candidate gene approach can be used when there is some pre-existing knowledge about the trait studied (Tabor et al. 2002). There are two main types of candidate genes, functional and positional. Functional candidates are genes with a putative role in disease development. Studies with animal models, association with other closely related diseases, information about pathways, and gene expression data can all increase understanding of the potential function of candidate genes (Cho and Weaver 2007). In contrast, positional candidate genes are identified through genome-wide scans. In this approach, genetic variations within the gene of interest are typed with polymorphic markers or haplotype tagging SNPs in the study population and the frequency of the variation is compared between patients and controls. Another possibility is to resequence the gene of interest in the study population in order to detect all variations within the gene. Variations can be divided into coding or non-coding SNPs, amino acid substitutions, small insertions or deletions, and copy number variations (CNVs). These variations can have an effect on the particular gene by altering the structure or expression of the encoded protein.

The most studied functional candidate genes in IBD susceptibility have been the human leukocyte antigen (HLA) molecules, based on their role in regulating the immune response.
and well-established association with other autoimmune diseases such as celiac disease, rheumatoid arthritis, and diabetes. TLR4 and tumor necrosis factor-\(\alpha\) (TNF-\(\alpha\)) as functional IBD candidate genes as well as genes of interleukin-1 (IL-1) family have also been investigated in the pathogenesis of IBD based on their role in immune response (Orchard et al. 2000). The most remarkable example of the positional candidate gene approach has been the identification of the CARD15 gene in CD susceptibility (Hampe et al. 2001; Ogura et al. 2001a).

3.2 Genome-wide linkage studies

A genome-wide linkage approach can be applied in families with multiple affected individuals when there is no previous information about the susceptibility genes. In the past, linkage studies have been successfully used to identify the genes underlying rare monogenic Mendelian disorders (Baron 2001; Hirschhorn and Daly 2005). The principle behind a genome-wide scan is to genotype highly informative markers, usually microsatellite markers, distributed over the whole genome and follow whether a certain marker is inherited together with a disease phenotype within a family. If the causal gene is located close enough to the marker it is likely to be co-inherited at meiosis. Markers that are closest to the disease gene show the strongest correlation, indicated as LOD (logarithm of odds) or NPL (non-parametric linkage) scores, with the disease phenotype in families. Genome-wide linkage analysis became feasible due to development of microsatellite marker maps and semi-automated genotyping techniques over the last decade. Between the years 1996-2004 all together 11 IBD genome-wide linkage scans were performed (Hugot et al. 1996; Satsangi et al. 1996a; Cho et al. 1998; Hampe et al. 1999a; Ma et al. 1999; Duerr et al. 2000; Rioux et al. 2000; Williams et al. 2002; Paavola-Sakki et al. 2003; Barmada et al. 2004; Vermeire et al. 2004). In addition, there is a meta-analysis using ten of the published genome-wide scans (van Heel et al. 2004).

To date there are nine IBD loci identified through linkage studies with varying degrees of replication and statistical support (Table 3). While some loci seem to be specific to CD or UC, others seem to confer susceptibility to IBD overall. The most consistent linkage finding is linkage between CD and the IBD1 locus in the pericentromeric region of chromosome 16, discovered in the first genome scan in 1996 (Hugot et al. 1996). The IBD1 locus in CD susceptibility has been replicated in many subsequent linkage studies (Ohmen et al. 1996; Brant et al. 1998; Cavanaugh et al. 1998; Curran et al. 1998; Annese et al. 1999; Hampe et al. 1999a; Cavanaugh 2001) with the greatest replication of linkage (MLOD 5.79) with 613 nuclear families provided by the IBD International Genetics Consortium (Cavanaugh 2001). It is presently known that the CARD15 gene in this region
confers susceptibility to CD, but CARD15 does not explain all the linkage seen at this locus. Linkages to the IBD3 and IBD5 loci are the next most consistent linkage findings in IBD genetics, and their role is described in more detail in section 5.

<table>
<thead>
<tr>
<th>LOCUS</th>
<th>CHROMOSOME</th>
<th>DISEASE</th>
<th>GENES</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBD4</td>
<td>14q11-q12</td>
<td>CD</td>
<td></td>
<td>Ma et al. 1999, Duerr et al. 2000, Vermeire et al. 2004</td>
</tr>
<tr>
<td>IBD5</td>
<td>5q31</td>
<td>CD</td>
<td>SLC22A4/5</td>
<td>Ma et al. 1999, Rioux et al. 2000, Williams et al. 2002</td>
</tr>
<tr>
<td>IBD6</td>
<td>19p13</td>
<td>IBD</td>
<td></td>
<td>Rioux et al. 2000, Paavola-Sakki et al. 2003</td>
</tr>
<tr>
<td>IBD7</td>
<td>1p36</td>
<td>IBD</td>
<td></td>
<td>Cho et al. 1998</td>
</tr>
<tr>
<td>IBD8</td>
<td>16p</td>
<td>CD</td>
<td></td>
<td>Hampe et al. 2002a*</td>
</tr>
<tr>
<td>IBD9</td>
<td>3p26</td>
<td>IBD</td>
<td></td>
<td>Satsangi et al. 1996a, Rioux et al. 2000</td>
</tr>
</tbody>
</table>

*Not a genome-wide linkage study

The first evidence for significant linkage in UC was detected in the second genome-wide scan (Satsangi et al. 1996a). This locus, named IBD2, was located on chromosome 12 with a significant LOD score of 5.47. This finding has later been replicated in many genome-wide scans (Hampe et al. 1999a; Ma et al. 1999; Paavola-Sakki et al. 2003; Barmada et al. 2004). Studies on the IBD2 locus have also shown linkage to both CD and UC (Duerr et al. 1998; Yang et al. 1999a; Parkes et al. 2000; Uthoff et al. 2002; Achkar et al. 2006). Despite several studies of candidate genes including interferon gamma (IFNG) (Hampe et al. 1998), vitamin D receptor (VDR) (Simmons et al. 2000; Martin et al. 2002), integrin beta-7 (ITGB7) (van Heel et al. 2001), and signal transducer and activator of transcription 6 (STAT6) (Xia et al. 2003), no susceptibility gene responsible for the IBD2 linkage has yet been identified.

3.2.1 Genome-wide linkage study in Finnish IBD families

A genome-wide scan in Finnish families with IBD provided evidence for novel susceptibility loci (Paavola-Sakki et al. 2003). This study was unique since the UC cohort
examined was the largest analysed from one population so far. Furthermore, it provided data on the genetically homogenous Finnish isolate which may have advantages in the mapping of complex disease genes. In the first phase of the study, 92 IBD families were genotyped for 429 microsatellite markers spaced at approximately 10 cM intervals. The second phase consisted of high density mapping of chromosomal regions 2p13-11, 11p12-q13, and 12p13-12 with 38 new families and 15 additional markers. The most significant two-point NPL scores for UC were 2.61 on chromosome 2p11 and 2.00 on 12p13. The highest two-point NPL score of 2.15 for IBD was also reached on the same 12p13 locus and the second best value of 2.0 was obtained on chromosome 11. For CD, the highest two-point NPL score was 2.34 on chromosome 12q23. In addition, a nominal multipoint NPL score of 2.11 was observed for chromosome 16. When CARD15 positive families were excluded, a multipoint NPL score of 1.57 was still observed at the same location. When the IBD cohort was stratified according to early age at onset, the linkage signals were further improved on chromosomes 2p11, 11p12-q13, 12p13, 19p13, and 19q13. The corresponding multipoint NPL scores were 2.34, 2.31, 2.19, 2.07, and 2.66, respectively. This study replicated the linkages at the IBD1, IBD2, and IBD6 loci (Table 3) and provided some evidence for novel loci on chromosomes 2, 11p12-q13, 12q23, and 19q13.

3.3 Genome-wide association studies

Genome-wide association (GWA) studies are becoming an increasingly popular tool for identifying complex disease genes with low or modest effects in large sample sets. The popularity of this method is due to advances in technology for high-throughput genetic analysis, with very low error rates and low genotyping costs (Amos 2007). Progress made by the Human Genome Project and the International HapMap Project have also helped to make GWA studies possible (Hirschhorn and Daly 2005; Cho and Weaver 2007). The study population in a GWA study consists of affected individuals and healthy controls which is generally much easier to obtain than a collection of families for linkage studies. The markers used in GWA studies are SNP markers and tests of association with the condition are conducted by comparing allele frequencies between patients and controls. All published GWA studies in IBD have so far focused on CD which is thought to be less genetically heterogeneous than UC (Yamazaki et al. 2005; Duerr et al. 2006; Consortium 2007; Hampe et al. 2007; Libioulle et al. 2007; Raelson et al. 2007; Rioux et al. 2007). Within a short period of time many novel candidate genes have been reported, at the moment the most replicated genes being IL23R and ATG16L1 (Table 4).
Table 4. Novel candidate genes and chromosomal loci identified by genome-wide association studies.

<table>
<thead>
<tr>
<th>GENE</th>
<th>dbSNP ID</th>
<th>CHROMOSOME</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNFSF15</td>
<td>several SNPs</td>
<td>9q33</td>
<td>Yamazaki et al. 2005</td>
</tr>
<tr>
<td>ATG16L1</td>
<td>rs2241880</td>
<td>2q37.1</td>
<td>Hampe et al. 2007, Libioulle et al. 2007, Rioux et al. 2007, Consortium 2007</td>
</tr>
<tr>
<td>PTGER4</td>
<td>several SNPs</td>
<td>5p13.1</td>
<td>Libioulle et al. 2007, Consortium 2007</td>
</tr>
<tr>
<td></td>
<td>rs224136</td>
<td>10q21.1</td>
<td>Rioux et al. 2007, Consortium 2007</td>
</tr>
<tr>
<td>PHOX2B</td>
<td>rs16853571</td>
<td>4p12</td>
<td>Rioux et al. 2007</td>
</tr>
<tr>
<td>NCF4</td>
<td>rs4821544</td>
<td>22q13.1</td>
<td>Rioux et al. 2007</td>
</tr>
<tr>
<td>FAM92B*</td>
<td>rs8059910</td>
<td>16q24.1</td>
<td>Rioux et al. 2007</td>
</tr>
<tr>
<td>IRGM</td>
<td>rs1000113</td>
<td>5q33.1</td>
<td>Consortium 2007</td>
</tr>
<tr>
<td>BSN, MST1</td>
<td>rs9858542</td>
<td>3p21</td>
<td>Consortium 2007, Raelson et al. 2007</td>
</tr>
<tr>
<td>NKX2-3</td>
<td>rs10883365</td>
<td>10q24.2</td>
<td>Consortium 2007</td>
</tr>
<tr>
<td>PTPN2</td>
<td>rs2542151</td>
<td>18p11</td>
<td>Consortium 2007</td>
</tr>
<tr>
<td>JAKMIP1, LOC285484</td>
<td>several SNPs</td>
<td>4p16.1</td>
<td>Raelson et al. 2007</td>
</tr>
<tr>
<td></td>
<td>rs4794986</td>
<td>17q11.1</td>
<td>Raelson et al. 2007</td>
</tr>
<tr>
<td></td>
<td>rs8077981</td>
<td>17q23.2</td>
<td>Raelson et al. 2007</td>
</tr>
</tbody>
</table>

* a predicted gene

4. IBD SUSCEPTIBILITY GENES

Over the years, several genes have been proposed in IBD susceptibility but they have demonstrated varying levels of replication by independent studies. Before GWA studies, CARD15, IBD5 (includes SLC22A4/5), and HLA class II associations represented the most consistent replications (Table 5). For many of the candidate genes, either inadequate or conflicting replication studies have been reported and therefore definite conclusions regarding IBD association can not be made. DLG5, ABCB1/MDR1, and TLR4 genes belong to this category of controversial results (Table 5). Two other candidate genes TNFRSF1A and TRAF6 were of special interest due to functional properties of the genes and previous results from the genome-wide linkage study of Finnish IBD families (Table 5). Based on the recent GWA studies, the most promising susceptibility genes are IL23R and ATG16L1 (Tables 4 and 5).

Even though there is still much to be discovered in the pathogenesis of IBD, genetic defects at different levels are suspected to lead to the development of chronic inflammation in IBD (Goyette et al. 2007). Mutations in environmental sensors, such as CARD15 and TLR4, and their signalling pathways including TRAF6 may alter the innate
# Table 5. IBD susceptibility and most promising candidate genes. Only positive associations are shown here.

<table>
<thead>
<tr>
<th>GENE</th>
<th>dbSNP ID</th>
<th>CHROMOSOME</th>
<th>DISEASE</th>
<th>POPULATION</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLC22A5</td>
<td>rs2631367</td>
<td>5q31</td>
<td>CD</td>
<td>Same as in SLC22A4</td>
<td>Same as in SLC22A4 without study of Silverberg et al. 2007</td>
</tr>
<tr>
<td>DLG5</td>
<td>DLG5_e26rs</td>
<td>10q23</td>
<td>IBD</td>
<td>British,</td>
<td>Stoll et al. 2004, Daly et al.</td>
</tr>
<tr>
<td>Gene/Region</td>
<td>Reference</td>
<td>Chromosome</td>
<td>Disorder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
<td>------------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCB1/MDR1</td>
<td>rs2032582, rs1045642, rs3789243</td>
<td>7q21.1</td>
<td>IBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>British, German, Italian, New Zealand, Scottish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TLR4</td>
<td>rs4986790, rs4986791</td>
<td>9q32-33</td>
<td>IBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Belgian, Dutch, German, Greek</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFRSF1A</td>
<td>rs767455</td>
<td>12p13.2</td>
<td>IBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Belgian, Canadian</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pierik et al. 2004, Waschke et al. 2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRAF6</td>
<td></td>
<td>11p12</td>
<td>IBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL23R</td>
<td>rs1004819, rs7517847, rs10489629, rs2201841, rs11465804, rs11209026, rs1343151, rs10889677, rs11209032, rs1495965</td>
<td>1p32.1-31.2</td>
<td>IBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dutch, English, German, Hungarian, Italian, New Zealand, North American, Scottish, Spanish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATG16L1</td>
<td>rs2241880</td>
<td>2q37.1</td>
<td>CD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>British, Dutch, German, New Zealand, North American, Scottish</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Immunity by causing a decrease in NF-κB activation and defensin production. DLG5 and ABCB1/MDR1 gene mutations may change the permeability of the epithelial barrier. In addition, mutations in the genes involved in adaptive immunity including HLA genes and TNFRSF1A as well as cytokine receptor IL23R may cause an imbalance between
regulatory and effector cell immune responses which are important in controlling inflammatory reactions.

4.1 CARD15: the first CD susceptibility gene

An initial link between chromosome 16 and CD was reported in 1996 (the IBD1 locus, Table 3) (Hugot et al. 1996). Five years later, the IBD International Genetics Consortium replicated the linkage on chromosome 16 (Cavanaugh 2001). Twelve groups from Europe, North America, and Australia provided 613 nuclear families which showed strong linkage to multiple markers on chromosome 16 with a peak LOD score of 4.96. When analysis was restricted to CD families only, a LOD score of 5.79 was detected in the same interval with no evidence of linkage in UC and mixed families. These results indicated that the gene(s) at the IBD1 locus likely conferred susceptibility solely for CD.

Later the same year, three independent groups reported that mutations in the caspase recruitment domain-containing protein 15 (CARD15) gene are associated with CD (Hampe et al. 2001; Hugot et al. 2001; Ogura et al. 2001a). This was a landmark discovery since CARD15 was the first susceptibility gene known to be involved in a complex disease and it also confirmed the previous epidemiological data suggesting that CD is a genetic disorder. The first two studies were reported independently in the same issue of Nature (Hugot et al. 2001; Ogura et al. 2001a) and the third study a month later in The Lancet (Hampe et al. 2001). Hugot and co-workers proceeded from the linkage observation to CARD15/NOD2 gene identification by positional-cloning strategy whereas the two other groups adopted CARD15 as a potential candidate gene. Three common variants, R702W, G908R, and 1007fs, were shown to associate with CD susceptibility, the 1007fs mutation being the most significant (Hugot et al. 2001). Analysing the results further it was observed that these three mutations were always on different haplotypes. The relative risk of developing CD in heterozygotes was estimated to be three-fold, and in homozygotes or compound heterozygotes about 40 fold. Ogura and co-workers, who had just identified the CARD15 gene (Ogura et al. 2001b), came to the same conclusion showing that the R702W, G908R, and 1007fs mutations are responsible for CD susceptibility (Ogura et al. 2001a). In the third study, the 1007fs mutation was shown to associate with CD in German and British cohorts (Hampe et al. 2001).

4.1.1 CARD15 structure, expression, and function

CARD15 encodes an intracellular nucleotide-binding oligomerization domain protein 2 (NOD2) receptor involved in regulating proinflammatory cytokines in response to bacteria
by inducing signalling pathways of nuclear factor-κB (NF-κB) and mitogen-activated protein kinases (MAPKs) (Akira and Takeda 2004; Strober et al. 2006; Cho and Abraham 2007; Kanneganti et al. 2007). NOD2 belongs to the NLR family that contains an N-terminal caspase recruitment domain (CARD), a central nucleotide binding domain (NBD also known as NACHT), and a C-terminal regulatory leucine-rich repeat (LRR) domain (Ogura et al. 2001b; Martinon and Tschopp 2005; Creagh and O'Neill 2006). The CARD domain is responsible for protein-protein interactions generating downstream signalling, the NBD domain mediates self-oligomerization and also has ATPase activity, and the LRR domain is involved in ligand recognition directly or indirectly (Strober et al. 2006; Bourhis et al. 2007). NOD2, being a pattern recognition receptor (PRR) in innate immunity, recognises muramyl dipeptide (MDP) in the cytoplasm (Bonen et al. 2003; Chamaillard et al. 2003; Girardin et al. 2003; Inohara et al. 2003). MDP is derived from the bacterial cell wall component peptidoglycan (PGN) which is a common component of both gram-positive and gram-negative bacteria.

When CARD15 was found, it was at first thought that its expression was restricted to peripheral blood monocytes (Ogura et al. 2001b). Later it was shown that NOD2 is also expressed in various other tissues such as dendritic cells (Gutierrez et al. 2002), endothelial cells (Davey et al. 2006), granulocytes (Gutierrez et al. 2002), intestinal myofibroblasts (Otte et al. 2003), Paneth cells of the small intestine (Lala et al. 2003; Ogura et al. 2003), and various epithelial cells including intestinal epithelium (Berrebi et al. 2003; Hisamatsu et al. 2003; Uehara et al. 2007).

4.1.2 NOD2 signalling pathway

Mechanisms by which muramyl dipeptide (MDP) enters the cytoplasm appear to be via a transporter protein hPepT1 (Vavricka et al. 2004; Ismair et al. 2006), by infection of invasive bacteria, through phagocytosis or secretion system-mediated translocation of muropeptides, and by host-cell mediated internalisation (Philpott and Viala 2004; Strober et al. 2006; Bourhis et al. 2007). Inside the cell MDP is recognised by the LRR domain of NOD2 (Figure 1). NOD2 has been shown to localise to cytoplasm, vesicles, and plasma membrane (Barnich et al. 2005; Lecine et al. 2007). The membrane association of NOD2 makes it possible to respond to MDP and activate the NF-κB pathway. NOD2 is also in charge of the membrane recruitment of a serine-threonine kinase RICK (also called RIP2, RIPK2 and CARDIAK), the next molecule in the pathway. RICK, in turn, binds directly to IkB kinase (IKK) γ (also known as NEMO) and activates the IKK complex by promoting the ubiquitinylation of IKKγ and by stimulating the kinase activity of IKKα and
Figure 1. NOD2 and TLR4 signalling pathways in antigen presenting cell (APC) resulting in production of pro- and anti-inflammatory cytokines. Possible interactions between these two pathways are indicated with a red arrow. NOD2 downstream signalling is mediated through a CARD domain and TLR4 signalling through a cytosolic TIR domain (Toll/interleukin-1 receptor). CD14=monocyte differentiation antigen CD14, IKK=I-kappa-B kinase, IRAK=interleukin 1 receptor-associated kinase, LBP=lipopolysaccharide-binding protein, LPS=lipopolysaccharide, MAPK=mitogen-activated protein kinase, MDP=muramyl dipeptide, MyD88=myeloid differentiation primary response gene 88, NOD2=nucleotide-binding oligomerization domain protein 2, NF-κB=nuclear factor-κB, P=phosphorylation, RICK=RIP-like interacting clarp kinase, TAB=TAK-binding protein, TAK1=transforming growth factor-beta-activated kinase 1, TLR4=toll-like receptor 4, TRAF6=TNF receptor-associated factor 6, Ub=ubiquitination.
IKKβ, the other two components of the IKK complex (Inohara et al. 2000; Abbott et al. 2004). Next, the phosphorylation of IKKβ and IkB leads to the release of NF-κB for translocation to the nucleus and production of proinflammatory cytokines. In addition to NF-κB, pathogen-induced expression and secretion of multiple cytokines is mediated through several transcription factors activated by MAPK-related kinases (Inohara et al. 2005). NOD2 induces activation of p38, JNK, and ERK but the exact molecular mechanism by which the activation is achieved is not well understood at the moment.

The NOD2 dependent signalling pathway is far from being fully characterised. NOD2 acts independently of the TLR cascade but it is likely that interplay between these two protein families ensures the efficient co-ordination of innate immune responses, through either synergistic or co-operative signalling (Figure 1) (Netea et al. 2004; Watanabe et al. 2004; Meinzer and Hugot 2005; van Heel et al. 2005b; Bourhis et al. 2007). One proposed model of interaction between these pathways is at the level of TAK1 as RICK has been shown to induce the activation of TAK1 (Windheim et al. 2007). The most recent study suggests that the signalling pathways converge at the level of TRAF6, which is known to be activated by NOD2 (Abbott et al. 2007).

4.1.3 CARD15 mutation frequency and genotype-phenotype associations in different CD cohorts

Investigations of the three common mutations R702W, G908R, and 1007fs in the CARD15 gene have demonstrated a notable amount of heterogeneity across different populations. Ethnically divergent populations may exhibit the same phenotype but this may not be due to the same set of variants, genes, or even loci. CARD15 variants are found in the majority of Caucasian CD patients (Table 5) although the mutation frequency in different populations is remarkably different. The overall frequencies of the common CARD15 mutations in North American and central European Caucasian CD populations vary between 27.0% and 46.1% (Hugot et al. 2001; Abreu et al. 2002; Ahmad et al. 2002; Cuthbert et al. 2002; Hampe et al. 2002b; Vermeire et al. 2002a; Vermeire et al. 2002b; Bairead et al. 2003; Fernandez et al. 2004; Giachino et al. 2004; Heresbach et al. 2004; Lakatos et al. 2004; Newman et al. 2004; Annese et al. 2005a; Lakatos et al. 2005; Bene et al. 2006; Ferraris et al. 2006; Oostenbrug et al. 2006b; Riis et al. 2007). In Scotland, which is a high incidence population, the allele frequencies were lower still (22.8%) as compared to other central European populations (Arnott et al. 2004). The mutation frequency in children with CD seems to be even higher, ranging from 38% to 67% (Sun et al. 2003; Bene et al. 2006; Cucchiara et al. 2007b). This is an indication for stronger genetic predisposition in early-onset CD.
REVIEW OF THE LITERATURE

The relative risk for CD when carrying one common CARD15 mutation is only about 2-4 fold whereas carriage of two mutations (compound heterozygote or homozygote) increases the relative risk dramatically from 15 to 40 fold (Hugot et al. 2001; Cuthbert et al. 2002; Economou et al. 2004; Newman et al. 2004; Vermeire and Rutgeerts 2005; Oostenbrug et al. 2006b; Henckaerts and Vermeire 2007).

Several studies have investigated the phenotypic expression of CARD15 variants, and the most consistently replicated association is of CARD15 mutations with ileal CD (Hampe et al. 2001; Abreu et al. 2002; Ahmad et al. 2002; Cuthbert et al. 2002; Hampe et al. 2002b; Lesage et al. 2002; Vermeire et al. 2002b; Bairead et al. 2003; Brant et al. 2003b; Arnott et al. 2004; Economou et al. 2004; Fernandez et al. 2004; Giachino et al. 2004; Heresbach et al. 2004; Lakatos et al. 2004; Annese et al. 2005a; Lakatos et al. 2005; Ferraris et al. 2006; Oostenbrug et al. 2006b). The stricturing and/or penetrating phenotype has also been associated with CARD15 mutations (Abreu et al. 2002; Hampe et al. 2002b; Lesage et al. 2002; Brant et al. 2003b; Economou et al. 2004; Giachino et al. 2004; Heresbach et al. 2004; Lakatos et al. 2005; Oostenbrug et al. 2006b; van der Linde et al. 2007). In some studies, carriage of two mutations has been shown to associate with an earlier age at onset (Lesage et al. 2002; Brant et al. 2003b; Newman et al. 2004; Annese et al. 2005a; Oostenbrug et al. 2006b). Conflicting results concerning the allele frequencies between familial and sporadic cases have been reported. In some studies the frequency has not been different (Newman et al. 2004; Annese et al. 2005a; van der Linde et al. 2007) whereas other studies have been able to show higher allele frequencies in familial patients (Cuthbert et al. 2002; Zhou et al. 2002; Economou et al. 2004; Oostenbrug et al. 2006b).

In Scandinavia, CD incidence is among the highest in the world. However, the allele frequencies of the three common CARD15 mutations have been reported to be lower than in central Europe and North America. Nevertheless, these variants have still been associated with CD as well as with specific phenotypes, namely earlier age at diagnosis (Ernst et al. 2007) and ileal disease location (Vind et al. 2005; Medici et al. 2006; Torkvist et al. 2006; Ernst et al. 2007). In Sweden, the common variants R702W and G908R, but not the 1007fs mutation, were associated with CD susceptibility (Torkvist et al. 2006). The contribution of the three mutations in Swedish CD patients was relatively low, with 15.2% of patients carrying one of the common mutations. CARD15 mutation frequency in Swedish pediatric patients with CD was even lower, only 8.6% (Idestrom et al. 2005). This may be partly due to a lower frequency of ileal disease in this cohort. Study of Norwegian population for the three mutations showed a low mutation frequency of 7.0% (Hampe et al. 2002b). In a second Norwegian study, the mutation frequency was similar with no significant CD association (Medici et al. 2006). Indeed, the R702W and
G908R variants had essentially identical allele frequencies among Norwegian controls and cases. The third 1007/fs had higher frequency in patients than in healthy controls (5.3% vs. 2.4%) but the difference was not statistically significant. In Denmark, approximately 21% of the CD patients studied carried at least one CARD15 mutation (Vind et al. 2005; Ernst et al. 2007). The R702W and 1007/fs mutations were also independently associated with CD in Danish patients. A recent publication reported a much higher CARD15 mutation frequency of 44% in Danish twins with CD (Jess et al. 2005). There is a single study concerning the CARD15 mutation frequency in Icelandic CD patients, where no one of 49 CD patients had any of the three common mutations (Thjodleifsson et al. 2003).

Jews have the highest incidence and prevalence rates of IBD compared to other ethnic groups. The mutation frequency of the CARD15 gene has been investigated in different subgroups of Jews, namely Ashkenazi and Sephardic populations. In North American Ashkenazi Jewish populations, 30-34% of CD patients carried at least one of the three common mutations (Zhou et al. 2002; Bonen et al. 2003; Newman et al. 2004; Tukel et al. 2004). In one study, patients carrying two mutations were shown to have an earlier age at onset, especially for familial patients (Zhou et al. 2002). CARD15 mutation frequency in Ashkenazi and Sephardic Jews from Israel has also been studied (Karban et al. 2004). The overall mutation frequency in CD patients was 41%, which is comparable to other Caucasian populations. A higher mutation frequency was observed in Ashkenazi compared to Sephardic Jews (48% vs. 28%). In another study, similar mutation frequency of 35% for a Sephardic Jewish population has been observed (Tukel et al. 2004).

The common CARD15 mutations are not a major contributor to CD in Asian populations. Two papers from Japan were published soon after the CARD15 finding and none of the main mutations was found in patient or control groups (Inoue et al. 2002; Yamazaki et al. 2002). Other negative reports on Asian populations involved Chinese (Leong et al. 2003; Gao et al. 2005) and South Korean (Croucher et al. 2003b; Lee et al. 2005) cohorts. These findings indicate that the three common CARD15 mutations are quite new and have arisen after the Asian-European population split. A significantly lower CARD15 mutation frequency has been identified in African-American compared to Caucasian cohorts (Bonen et al. 2002). Little is also known about African and South American populations with one exception of low mutation frequency in coloured South Africans (Zaahl et al. 2005).

4.1.4 Rare CARD15 variants and their effect on phenotype

On average, 60-70% of CD patients do not carry any of the three CARD15 mutations (Abreu et al. 2002; Ahmad et al. 2002; Cuthbert et al. 2002; Hampe et al. 2002a;
Vermeire et al. 2002a). These mutations do not account for all the linkage observed between CD and the IBD1 locus (Hugot et al. 2001; Hampe et al. 2002a; Paavola-Sakki et al. 2003). Since the residual linkage at the IBD1 locus was detected after removing the families with three variants, a search for additional disease-causing mutations in the coding region of the CARD15 gene has also been conducted. The first comprehensive study comprised 453 CD patients (Lesage et al. 2002). The authors identified 67 sequence variants of which 27 were rare and considered as potential disease-causing mutations (Table 6). Three common mutations accounted for 81% of the CD associated CARD15 variants, whereas the remaining 19% consisted of rare mutations. In addition, 50% of CD patients carried one mutation and 17% had two mutations, confirming the gene-dosage effect. Patients with two mutations more frequently had the stricturing phenotype as compared to patients without mutation (53% vs. 28%, p=0.00003). Double mutant patients had also less colonic involvement when compared to wild type patients (43% vs. 62%, p=0.003). When comparing age at onset between these two groups, double mutation carriers had slightly lower age at onset (16.9 vs. 19.8 years, p=0.01).

<table>
<thead>
<tr>
<th>Nucleotide change</th>
<th>Amino acid change</th>
<th>Protein domain</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exon 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G337A</td>
<td>D113N</td>
<td>CARD1</td>
<td>Tukel et al. 2004</td>
</tr>
<tr>
<td>G413A</td>
<td>R138Q</td>
<td>CARD2</td>
<td>Lesage et al. 2002</td>
</tr>
<tr>
<td>G418A</td>
<td>A140T</td>
<td>CARD2</td>
<td>Lesage et al. 2002</td>
</tr>
<tr>
<td>T469C</td>
<td>W157R</td>
<td>CARD2</td>
<td>Lesage et al. 2002</td>
</tr>
<tr>
<td><strong>Exon 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C566T</td>
<td>T189M</td>
<td>CARD2</td>
<td>Lesage et al. 2002</td>
</tr>
<tr>
<td><strong>Exon 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C703T</td>
<td>R235C</td>
<td></td>
<td>Lesage et al. 2002, King et al. 2006</td>
</tr>
<tr>
<td>G871A</td>
<td>D291N</td>
<td>NBD</td>
<td>Lesage et al. 2002</td>
</tr>
<tr>
<td>C931T</td>
<td>R311W</td>
<td>NBD</td>
<td>Lesage et al. 2002, Schnitzler et al. 2006</td>
</tr>
<tr>
<td>A1070C</td>
<td>D357A</td>
<td>NBD</td>
<td>Tukel et al. 2004</td>
</tr>
<tr>
<td>A1087T</td>
<td>I363F</td>
<td>NBD</td>
<td>Tukel et al. 2004</td>
</tr>
<tr>
<td>C1117T</td>
<td>R373C</td>
<td>NBD</td>
<td>Lesage et al. 2002</td>
</tr>
<tr>
<td>C1171T</td>
<td>R391C</td>
<td>NBD</td>
<td>Schnitzler et al. 2006</td>
</tr>
<tr>
<td>G1178A</td>
<td>R393H</td>
<td>NBD</td>
<td>King et al. 2006</td>
</tr>
<tr>
<td>A1241G</td>
<td>N414S</td>
<td>NBD</td>
<td>Lesage et al. 2002</td>
</tr>
<tr>
<td>Chromosome</td>
<td>Variant</td>
<td>Description</td>
<td>References</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>C1292T</td>
<td>S431L</td>
<td>NBD</td>
<td>Lesage et al. 2002, King et al. 2006</td>
</tr>
<tr>
<td>G1321A</td>
<td>E441K</td>
<td>NBD</td>
<td>Lesage et al. 2002, Tukel et al. 2004</td>
</tr>
<tr>
<td>C1387G</td>
<td>P463A</td>
<td>NBD</td>
<td>Schnitzler et al. 2006</td>
</tr>
<tr>
<td>C1648G</td>
<td>L550V</td>
<td>NBD</td>
<td>Tukel et al. 2004</td>
</tr>
<tr>
<td>1671delCCTGGG</td>
<td>558delLG</td>
<td>NBD</td>
<td>Lesage et al. 2002</td>
</tr>
<tr>
<td>C2003T</td>
<td>P668L</td>
<td></td>
<td>van Heel et al. 2006</td>
</tr>
<tr>
<td>C2050T</td>
<td>R684W</td>
<td></td>
<td>Lesage et al. 2002</td>
</tr>
<tr>
<td>C2137T</td>
<td>R713C</td>
<td></td>
<td>Lesage et al. 2002, Lakatos et al. 2005</td>
</tr>
<tr>
<td>G2138A</td>
<td>R713H</td>
<td></td>
<td>Schnitzler et al. 2006</td>
</tr>
<tr>
<td>C2278T</td>
<td>R760C</td>
<td>LRR1</td>
<td>Schnitzler et al. 2006</td>
</tr>
<tr>
<td>C2368T</td>
<td>R790W</td>
<td>LRR2</td>
<td>Schnitzler et al. 2006</td>
</tr>
<tr>
<td>C2372A</td>
<td>R791Q</td>
<td>LRR2</td>
<td>Sugimura et al. 2003a, Lakatos et al. 2005</td>
</tr>
<tr>
<td>C2371T</td>
<td>R791W</td>
<td>LRR2</td>
<td>Schnitzler et al. 2006</td>
</tr>
</tbody>
</table>

**Exon 5**

| C2475G    | N825K  | LRR4        | Schnitzler et al. 2006 |
| G2489A    | R830Q  | LRR4        | King et al. 2006, van Heel et al. 2006 |
| G2527A    | E843K  | LRR4        | Lesage et al. 2002 |
| C2546T    | A849V  | LRR4        | Schnitzler et al. 2006 |

**Exon 6**

| A2555G    | N852S  | LRR4        | Tukel et al. 2004, King et al. 2006 |
| A2558G    | N853S  | LRR5        | Lesage et al. 2002 |

**Exon 7**

| G2656A    | A885T  | LRR6        | Lesage et al. 2002 |
| C2686T    | R896X  | LRR6        | King et al. 2006, van Heel et al. 2006 |
Pediatric CD patients (n=55) with average age of 11.2 at onset were studied for additional \textit{CARD15} mutations and three potential disease-causing mutations were identified (R703C, E778K, and V793M, Table 6) (Sun et al. 2003). 67% of pediatric CD patients carried at least one of the three common mutations and the frequencies of other three missense mutations R703C, E778K, and V793M were 2% each. Nineteen (35%) patients were homozygous or compound heterozygous for the six disease-associated mutations and the 1007\textit{fs} mutation was significantly more common in pediatric CD patients than in healthy controls (allele frequency 26% vs. 2%). Genotype-phenotype correlations showed a significant association between \textit{CARD15} mutations and ileocolonic disease. Patients with two mutations had also more frequently undergone intestinal resection indicating possibility of more severe disease (=stricturing behaviour). All disease-associated alleles were in or close to the LRR domain which supports the finding of Lesage and co-workers (Lesage et al. 2002). The much higher frequency of double mutation carriers among the pediatric group also supports the idea of a gene-dosage effect for earlier age at onset.

In one study, 527 CD patients were screened for additional \textit{CARD15} mutations in exon 4 in addition to the common mutations (Lakatos et al. 2005). As a result, five rare variants (R713C, A755V, E778K, R791Q, and V793M) were found and the R703C variant was identified as a risk factor for CD (2.1% vs. 0% controls, p=0.02) (Table 6). In other two studies, additional \textit{CARD15} mutations were sought in a study population consisting of patients heterozygous for common \textit{CARD15} mutations (King et al. 2006; Schnitzler et al. 2006). As a result of screening the coding region, 11 of 100 CD patients had a second potential disease-causing mutation (R235C, R393H, S431L, A612T, R703C, V793M, R830Q, N852S, M863V, R896X, and R1019G, Table 6) (King et al. 2006). In another study, nine rare potential disease-causing mutations were found, eight of them novel (R391C, P463A, R713H, R760C, R790W, R791W, N825K, and A849V) and one described previously (N289S) (Table 6) (Schnitzler et al. 2006). Rare variants R311W,
R703C, and A755V (Table 6) detected were considered as polymorphisms. A severe phenotype was observed in CD patients who were compound heterozygotes for the common and the novel CARD15 mutation. In addition, a CARD15 functional study reported an additional rare variant in seven heterozygous CD patients (P668L, R703C, A755V, R830Q, R896X, E977K, and R1019G, Table 6) (van Heel et al. 2006).

Although CD is more common among Jews (Roth et al. 1989), the frequency of the three common mutations are similar in Jews and Caucasians. Initial results using six microsatellite markers spanning the IBD1 locus showed that Ashkenazi Jews might have additional mutations in this region (Sugimura et al. 2003a). Further evidence for additional variants was seen in a highly significant association between the SNP5 haplotype and Jewish CD patients. This led to the identification of two potential disease-associated alleles IVS8+158C→T and R791Q (Table 6). A haplotype carrying this intronic variant showed a higher risk for CD in Ashkenazi Jews than in non-Jews, even though the function of this haplotype remained unrecognised. In the second Jewish study, CARD15 screening revealed eight potential disease-causing variants D113N, L248R, D357A, I363F, L550V, A612T, R703C, and N852S (Table 6) (Tukel et al. 2004).

Due to the absence of the three CARD15 mutations in Japanese CD patients, screening of the whole coding area has been performed (Sugimura et al. 2003b). No amino acid change predisposing to CD was observed, further confirming that CARD15 is not a contributor for CD susceptibility in Japan.

4.1.5 Functional consequences of CARD15 mutations

NOD2-mediated functions identified to date include induction of pro- and anti-inflammatory cytokines and epithelial cell defences, up-regulation of costimulatory molecules leading to enhanced T cell activation, and mediation of antibacterial effect through different mechanisms (Cho and Abraham 2007). Defects in NOD2 function most likely contribute to some extent to alterations in intestinal immune homeostasis but how CD associated CARD15 mutations increase the susceptibility to CD remains unknown. It has been shown that CARD15 mutations lead to both gain and loss of function of this gene.

Transient transfection of cells has shown that the three common CARD15 mutations result in defective recognition of MDP and this further leads to decreased NF-κB activation (Bonen et al. 2003; Chamaillard et al. 2003; Girardin et al. 2003; Inohara et al. 2003). In addition to three common CARD15 mutations, multiple rare variants have been deficient
in their ability to sense MDP (Chamaillard et al. 2003). In another study, a comprehensive library of \textit{CARD15} variants was generated resulting in 519 amino acid changes and their function was investigated in HEK293 cells (Tanabe et al. 2004). Essential regulatory domains and specific residues of \textit{CARD15} involved in MDP recognition were identified. Amino acid residues 855-1040 in the LRR domain of the protein were found to be essential for the response to MDP.

A more sophisticated method for investigating NOD2 function has been the use of mononuclear cells derived from CD patients. In these studies the functional consequences of \textit{CARD15} in response to MDP has been determined by the production of different cytokines. The 1007\textit{fs} mutation results in defective release of anti-inflammatory cytokine IL-10 from blood mononuclear cells of CD patients after stimulation with the TLR2 ligands peptidoglycan and Pam3Cys-KKKK (Netea et al. 2004). In other study the 1007\textit{fs} mutation was reported to exhibit defects in production of proinflammatory cytokines TNF-\textit{\alpha}, IL-6, and IL-8 (van Heel et al. 2005a). Compound heterozygotes of the three common mutations also showed reduced IL-8 production when compared to wild types. CD patients homozygous for the 1007\textit{fs} mutation have also exhibited strongly decreased IL-1\textit{\beta} synthesis in response to both PGN and MDP (Netea et al. 2005). Later it has been shown that healthy individuals heterozygous for the common \textit{CARD15} mutation also have modest impairment of IL-8 response (van Heel et al. 2006). Macrophages from homozygous 1007\textit{fs} patients have demonstrated a loss of NOD2-induced signalling in a microarray study (Li et al. 2004). There is now compelling evidence from primary cell studies to suggest that the common \textit{CARD15} mutations lead to loss of function. However, this is inconsistent with the fact that CD is associated with overexpression of NF-\kappaB target genes TNF-\textit{\alpha}, IL-1\textit{\beta}, IL-6, and IL-12 under inflamed conditions (Monteleone et al. 1997; Parronchi et al. 1997; Rogler et al. 1998; Schreiber et al. 1998).

Studies of \textit{CARD15} knockout mice have generated conflicting data as to whether mutations in this gene cause loss or gain of function (Watanabe et al. 2004; Kobayashi et al. 2005; Maeda et al. 2005). Using NOD2 knockout mice, Watanabe and colleagues showed that NOD2 signalling in normal circumstances inhibits the TLR-driven T helper 1 response by regulating NF-\kappaB signalling (Watanabe et al. 2004). Thus, in the absence of NOD2 signalling there is no NOD2 inhibition and this results in increased TLR2-mediated NF-\kappaB activation and even more IL-12 production. In another study, a mouse model of mutant NOD2 has exhibited elevated NF-\kappaB activation in response to MDP and more efficient processing and secretion of the cytokine IL-1\textit{\beta} (Maeda et al. 2005). However, in a third study on NOD2 knockout mice, protective immunity was abolished as a
consequence of loss of function of NOD2 (Kobayashi et al. 2005). It is still unclear why CARD15 mutations seem to result in loss of function in humans and gain of function in mouse models and additional studies are needed to reconcile the differences between human and murine findings. NOD2 knockout mice do not spontaneously develop ileitis (Pauleau and Murray 2003; Kobayashi et al. 2005), demonstrating the requirement for other genetic factors for disease expression.

Studies on CARD15 mutants, both in humans and mice, have demonstrated that normal gene function is required for optimal defensin expression (Wehkamp et al. 2004; Kobayashi et al. 2005; Wehkamp et al. 2005). Defensins are antimicrobial peptides which are synthesized and secreted by Paneth cells of the small intestine (Vignal et al. 2007). Antimicrobial peptides of the epithelia are known to kill a broad range of microorganisms, thus protecting the host from pathogens. Ileal expression of human α-defensin 5 (HD-5) and HD-6 is decreased in the affected ileum of CD patients and more importantly, a significant decrease in expression was observed in CD patients with CARD15 mutations (Wehkamp et al. 2004; Wehkamp et al. 2005). Consistently, NOD2 knockout mice have a selective decrease in α-defensin expression (Kobayashi et al. 2005).

4.1.6 CARD15 in Blau syndrome and early-onset sarcoidosis

The CARD15 gene is also associated with Blau syndrome (BS; MIM 186580), a rare autosomal dominant disorder characterised by early-onset granulomatous arthritis, uveitis, and skin rash with camptodactyly (Blau 1985). Before identification of CARD15 in BS susceptibility, the susceptibility locus had already been mapped to chromosome 16p12-q21 by linkage analysis (Tromp et al. 1996). Both CD and BS are associated with granulomatous inflammation but the disease-causing mutations in the CARD15 gene are different. Four missense mutations (R334Q, R334W, E383K, and L469F) associated with BS are located in the central nucleotide-binding domain (NBD) (Miceli-Richard et al. 2001; van Duist et al. 2005) and exhibit enhanced basal NF-κB activity in contrast to CD associated mutations which cause decreased NF-κB activation in vitro (Chamaillard et al. 2003). However, a functional study on R334W mutation using patient mononuclear cells has shown normal responses to MDP (van Heel et al. 2006).

BS shows phenotypic overlap with early-onset sarcoidosis (EOS; MIM 609464). EOS is a rare disease usually starting before four years of age with a combination of skin, joint, and eye disorders (Kanazawa et al. 2005). EOS is also associated with mutations in the CARD15 gene and the BS associated mutation R334W is also found in EOS (Kanazawa et al. 2004; Kanazawa et al. 2005; Coto-Segura et al. 2007). In addition, D382E, H496L,
M513T, T605P, A612T, and N670K missense mutations have been found in EOS patients and their functional consequences as well as the R334W mutation have been examined in transfected HEK293T cells (Kanazawa et al. 2005). All except the A612T variant significantly increased basal NF-κB activity when compared to the wild type. However, MDP-induced NF-κB activity was not greater for mutants than for the wild type.

4.2 The DLG5 gene on chromosome 10

In 2004, a positional cloning approach led to identification of the discs large, drosophila, homolog of 5 (DLG5) gene (Table 5) (Stoll et al. 2004) after initial linkage of IBD to the pericentromeric region of chromosome 10 in a European cohort (Hampe et al. 1999a). DLG5 is a member of the MAGUK (membrane-associated ganylate kinase) gene family which encodes scaffolding proteins involved in intracellular signal transduction (Gonzalez-Mariscal et al. 2000). DLG5, with multiple protein interaction domains, has also been implicated in maintaining epithelial integrity and polarity (Nakamura et al. 1998; Humbert et al. 2003). It is also widely expressed in human tissues as well as the GI tract (small intestine, intestinal epithelial cells, and colon) (Ferguson et al. 2007).

The association of genetic variations in the DLG5 gene with IBD and particularly CD was initially described in German and UK study samples with two distinct haplotypes (Stoll et al. 2004). The haplotype A (protective), tagged by SNP DLG5_e26 ins/delA, was significantly undertransmitted in IBD and CD, whereas the haplotype D (risk), tagged by the SNP R30Q (G89A, previously reported as G113A), was significantly overtransmitted in these groups. The R30Q polymorphism is located in the DUF622 domain of the DLG5 gene and is likely to have functional consequences. Further studies have, however, yielded inconsistent results (Table 5) (Yamazaki et al. 2004; Daly et al. 2005; Gazouli et al. 2005a; Noble et al. 2005b; Torok et al. 2005; Vermeire et al. 2005; Buning et al. 2006; Lakatos et al. 2006; Medici et al. 2006; Newman et al. 2006; Tremelling et al. 2006; Browning et al. 2007; Pearce et al. 2007; Russell et al. 2007), indicating that the effect appears to be inconsistent, relatively weak, and population specific (Russell et al. 2007). Phenotypic association between the rare variant 30Q and ileal CD has been reported (Pearce et al. 2007). Meta-analysis gave no evidence of association of 30Q with IBD indicating that this variant is not a risk factor for IBD across populations (Browning et al. 2007). In addition, two studies have reported that the rare variant is a risk factor for CD in men both in adults and children (Friedrichs et al. 2006; Russell et al. 2007) whereas one study on pediatric CD proposes that this variant has a protective effect in girls (Biank et al. 2007). The meta-analysis concerning the gender-stratified analysis from 11 published
studies showed that DLG5 R30Q is associated with a small decrease in risk of CD in women (Browning et al. 2008).

4.3 The ABCB1/MDR1 gene on chromosome 7

The ATP-binding cassette, sub-family B, member 1 (ABCB1) gene (also known as the MDR1 gene) is located on chromosome 7q21, in a region of suggestive evidence for linkage for both CD and UC (Satsangi et al. 1996a; Cho et al. 1998; van Heel et al. 2004). ABCB1 knockout mice also spontaneously develop colitis in a specific pathogen-free environment (Panwala et al. 1998). The condition can be prevented by and treated with antibiotics, suggesting that intestinal flora are necessary to initiate and maintain the inflammation (Annese et al. 2006). The ABCB1 gene encodes P-glycoprotein 170, which functions as a transmembrane ATP-dependent efflux pump moving drugs and toxins from the intracellular to the extracellular domain (Annese et al. 2006). The protein is highly expressed at the apical surface of epithelia of the colon and distal small intestine (Thiebaut et al. 1987) and this suggests a role in protection not only against xenobiotics but also bacterial products (Ho et al. 2005). The protein has also been detected in haematopoietic stem cells, macrophages, dendritic cells, natural killer cells, and both T and B lymphocytes (Klimecki et al. 1994). ABCB1 expression is significantly decreased in the colon of UC but not CD patients according to a recent finding (Langmann et al. 2004).

Two exonic polymorphisms, C3435T and G2677T/A (A893S/T), have received the most attention. Although the C3435T SNP does not change the amino acid, it has been shown to alter protein expression (Hoffmeyer et al. 2000). The other variant may also have an influence on protein expression (Kim et al. 2001; Tanabe et al. 2001). Association with CD and/or UC have been established in some, but not all, studied populations (Table 5) (Brant et al. 2003a; Croucher et al. 2003a; Schwab et al. 2003; Gazouli et al. 2004; Glas et al. 2004; Potocnik et al. 2004; Ho et al. 2005; Palmieri et al. 2005; Onnie et al. 2006b; Oostenbrug et al. 2006a; Osuga et al. 2006; Urcelay et al. 2006; Ardizzone et al. 2007; Cucchiara et al. 2007a; Farnood et al. 2007). A meta-analysis of nine published studies of the C3435T variant in IBD showed a significant association between the rare T allele and UC (Onnie et al. 2006b). In another meta-analysis, which partly employed the same published studies, an equivalent conclusion was made (Annese et al. 2006).

In addition, a gene-wide haplotype tagging strategy was recently conducted using six haplotype tagging SNPs (tSNPs) and showed a significant association especially with extensive UC which was critically dependent on one tSNP, an intronic variant rs3789243
(Ho et al. 2006). All haplotypes carrying this intronic variant maintained a highly significant association and the effect of this tSNP was independent of the C3435T variant.

4.4 The TLR4 gene on chromosome 9

The innate immune response is the first line of defence against microbial infections. Together with the NLR family, the toll-like receptor (TLR) family proteins are in charge of the detection of different bacterial and viral components (Doyle and O’Neill 2006). These two protein families, known as pathogen-recognition receptors (PRRs), appear to play essential roles in mucosal homeostasis and alterations contributing to the pathogenesis of IBD (Yamamoto-Furusho and Podolsky 2007). Identification of CARD15 as a CD susceptibility gene has raised the question whether other PRRs are involved in the pathogenesis of IBD. Toll-like receptor 4 (TLR4), in combination with CD14, LBP, and MD-2 proteins, acts as the PRR for lipopolysaccharide (LPS), a cell wall component of gram-negative bacteria (Aderem and Ulevitch 2000; Barton and Medzhitov 2003). Recognition of LPS triggers a cascade of cellular signals that culminate in NF-κB activation and, further, leads to proinflammatory cytokine production and clearance of the infectious agent (Figure 1) (Doyle and O’Neill 2006).

In mammals, the intestinal epithelium is a major barrier to the entry of microbes into an organism. TLR4 is expressed on the cell surface of antigen presenting cells including macrophages and dendritic cells as well as in the intestinal epithelium. Furthermore, alterations of TLR4 expression by intestinal epithelial cells have been described in IBD, showing that normally TLR4 is only barely detectable whereas in both UC and CD patients TLR4 is strongly upregulated (Cario and Podolsky 2000). Increased expression of TLR4 by myeloid dendritic cells in IBD has also been reported (Hart et al. 2005).

Two common polymorphisms (D299G and T399I) in the extracellular domain of the TLR4 gene have been described in humans. The D299G polymorphism has been associated with decreased responsiveness followed by LPS stimulation (Arbour et al. 2000; Schwartz 2002; Michel et al. 2003) although recent data have shown a normal response (Erridge et al. 2003; Heesen et al. 2003; Imahara et al. 2005). The potential association between IBD and these two genetic variations in TLR4 has been investigated by many groups (Table 5) (Arnott et al. 2004; Franchimont et al. 2004; Torok et al. 2004; Braat et al. 2005; Brand et al. 2005; Gazouli et al. 2005b; Lakatos et al. 2005; Oostenbrug et al. 2005; Ouburg et al. 2005; Riis et al. 2007). The results have been contradictory, which has been explained to be due to either population-related genetic heterogeneity, stratification bias, or sample size (Arnott et al. 2005). The most recent study, combining results of its own and eight
published reports concerning the D299G variant and a total of 4805 IBD patients, showed an association with IBD and also with CD (De Jager et al. 2007).

4.5 The TRAF6 gene on chromosome 11

Regulation of intestinal innate immunity involves cell surface and intracellular PRRs, adapter proteins, kinases, and the NF-κB system. Tumor necrosis factor (TNF) receptor-associated factors constitute a family of signal transducers that are thought to play important functions in innate and adaptive immunity, inflammation, and tissue homeostasis (Wu and Arron 2003). A member of this family, TRAF6, is an adaptor protein in the transcription factor NF-kB and activator protein-1 (AP-1) pathways mediating signalling downstream from receptors of the TNFR superfamily (though not TNFRSF1A) and TLR/IL-1R family (including TLR4) (Figure 1) (Cao et al. 1996; Deng et al. 2000; Ye et al. 2002; Wu and Arron 2003). While NF-κB is known to promote expression of genes involved in inflammatory responses and protection from apoptosis (Beg and Baltimore 1996), the stimulation of AP-1 activity may elicit stress responses and promote both cell survival and death (Shaulian and Karin 2001). In addition, the most recent findings suggest that NOD2 activation can cause the activation of TRAF6, making TRAF6 a common factor in two PRR pathways (Figure 1) (Abbott et al. 2007). Interestingly, one of the nominal linkages in the Finnish IBD scan was localised to chromosomal region 11p12-q13, where TRAF6 is also located (Paavola-Sakki et al. 2003).

4.6 The TNFRSF1A gene on chromosome 12

Tumor necrosis factor-alpha (TNFα) is a proinflammatory cytokine mediating apoptosis, inflammation, and immunity, and has been implicated in the pathogenesis of several human diseases, including IBD (Chen and Goeddel 2002). TNFα is secreted primarily by activated monocytes, macrophages, and T cells where activation of TLRs by a variety of PAMPs or cytokine receptors such as IL-1R stimulates transcription factor NF-κB that increases TNFα expression (Shealy and Visvanathan 2008). TNFα, in turn, signals through two distinct cell surface receptors TNFRSF1A and TNFRSF1B. TNF receptor superfamily member 1A (TNFRSF1A) (also known as TNFR1 and p55) is expressed on virtually all nucleated cells while the expression of TNF receptor superfamily member 1B (TNFRSF1B) is limited to immune and endothelial cells (Aggarwal 2003). TNFα/TNF receptor interactions play an essential role in the pathogenesis of the inflammatory response.
It has been shown that TNFα levels are elevated in serum, stool, and inflamed bowel mucosa of IBD patients (Braegger et al. 1992; Murch et al. 1993; Nicholls et al. 1993; Breese et al. 1994; Komatsu et al. 2001; Borruel et al. 2002). Several animal models have provided the foundation for the development of anti-TNF antibodies as treatment for chronic inflammatory diseases. IgG1 monoclonal antibodies against TNFα are an important advance in the treatment of CD and lately also of UC. The goal of this treatment is to reduce TNFα levels in the circulation to improve the clinical signs of the disease without causing systemic immunosuppression (Shealy and Visvanathan 2008).

Given the importance of TNF-receptor mediated signalling pathway in IBD and, in addition, that the TNFRSF1A gene is located near the IBD2 locus, the association of TNFRSF1A to IBD has also been studied (Table 5). In the first such study, an association between the A36G variant of the TNFRSF1A gene and pancolitis in UC was observed (Pierik et al. 2004). In another study, there was an association of the same TNFRSF1A gene variant with CD (Waschke et al. 2005). The same variant was negatively associated with the stricturing CD phenotype. However, in Japan no association between the A36G variant and IBD has been reported (Sashio et al. 2002).

4.7 The IL23R gene on chromosome 1

Recently, a highly significant association between ileal CD and the IL23R gene on chromosome 1p31 was reported (Table 4) (Duerr et al. 2006). In this genome-wide association study over 300 000 SNP markers were typed in a North American ileal CD case-control material. The most significant associations were observed to the CD susceptibility gene CARD15 and to the IL23R gene. A nonsynonymous IL23R SNP (R381Q) conferred a strong protection against CD. Nine additional noncoding markers at this locus were also associated. In the same study, replication analyses on a Jewish case-control cohort as well as 883 nuclear families were positive. The latter analysis also showed association with non-Jewish UC. Three subsequent genome-wide association studies have also identified the IL23R gene in CD susceptibility (Consortium 2007; Libioulle et al. 2007; Raelson et al. 2007).

Several replication studies on IL23R have already been conducted and association with both CD and UC has been reported (Table 5) (Borgiani et al. 2007; Buning et al. 2007; Cummings et al. 2007a; Glas et al. 2007; Oliver et al. 2007; Roberts et al. 2007; Tremelling et al. 2007; Weersma et al. 2008). The protective R381Q allele has been reported in pediatric CD (Baldassano et al. 2007b; Dubinsky et al. 2007) and IBD (Van
Limbergen et al. 2007a) as well. The only negative association was with Japanese CD patients (Yamazaki et al. 2007).

The IL23R protein was identified and characterised in 2002 (Parham et al. 2002). The functional IL23R protein is heterodimeric, comprising the IL23R subunit on chromosome 1p31 and an IL12Rβ1 subunit on chromosome 19p13. The latter subunit is also common to the IL12 receptor which has an IL12Rβ2 subunit as the second part. The ligand, a proinflammatory cytokine interleukin 23 (IL-23), is also heterodimeric and is composed of a unique p19 subunit (chromosome 12q13) and p40 subunit (chromosome 5q33), which is also a component of IL-12 (Oppmann et al. 2000). The other component of IL-12 is the p35 subunit. IL-23 is a pivotal cytokine in the differentiation of T helper cells, especially their differentiation into Th17 T cells (Neurath 2007; Steinman 2007). In IBD, the well controlled balance of the intestinal immune system is disturbed and myeloid dendritic cells falsely recognise commensal bacteria as pathogens (Baumgart and Carding 2007). This leads to maturation of dendritic cells and they further promote differentiation of naive T cells (Th0) into effector Th1, Th2, and Th17 cells as well as natural killer T cells. Traditionally, it has been thought that CD is driven through Th1 cell response and UC through Th2 cells (Baumgart and Carding 2007). Th17 cells on the other hand are different from classic Th1 and Th2 cells, both in the cytokines that induce their differentiation and the cytokines they produce (Cho and Weaver 2007). Clear evidence exists now for the central role of IL-23 in the development of CD.

4.8 The ATG16L1 gene on chromosome 2

In the European GWA study of nearly 20,000 nonsynonymous SNPs, an amino acid substitution (T300A) within the autophagy related 16-like 1 (ATG16L1) gene was found to be highly associated with CD (Table 4) (Hampe et al. 2007). Haplotype and regression analyses showed that the coding variant T300A carried virtually all the disease risk for CD exerted by this locus. Soon after that, three other GWA studies reported the same finding (Consortium 2007; Libioulle et al. 2007; Rioux et al. 2007) and several replication studies have also confirmed the association with CD (Table 5) (Baldassano et al. 2007a; Cummings et al. 2007b; Prescott et al. 2007; Roberts et al. 2007; Glas et al. 2008; Van Limbergen et al. 2008; Weersma et al. 2008). Two studies have reported phenotypic association of the ATG16L1 gene variation with ileal CD (Prescott et al. 2007; Van Limbergen et al. 2008). A modest association with UC has also been demonstrated in one study (Prescott et al. 2007).
The ATG16L1 protein is involved in the autophagocytosis pathway, which is known to account for degradation of long-lived proteins and organelles and in addition to mediate resistance to intracellular pathogens such as bacteria and viral particles (Schmid et al. 2006). Autophagic degradation links innate with adaptive immunity by delivering antigens to MHC class II presentation. ATG16L1 is one component of a large protein complex essential for autophagy (Mizushima et al. 2003). Expression of ATG16L1 has been detected in various tissues including the colon, small intestine, intestinal epithelial cells, leukocytes (T and B cells), and spleen (Hampe et al. 2007; Rioux et al. 2007). RNA-silencing methods have been used to knock-down the expression of the gene in human HeLa cells, and this decreased autophagy of Salmonella typhimurium (Rioux et al. 2007).

5. STUDIES OF SPECIFIC GENE LOCI

5.1 Chromosome 5 risk haplotype (IBD5 locus)

The IBD5 locus on chromosome 5 was originally identified in 1999 (Table 3) (Ma et al. 1999). This finding was replicated by higher-density mapping a year later in a Canadian CD cohort (Rioux et al. 2000) and the same authors further narrowed the associated area to chromosome 5q31 (Rioux et al. 2001). In this region a 250 kilobase (kb) haplotype shows strong association with CD. The haplotype contains 11 haplotype tagging SNPs equally associating with CD which means strong linkage disequilibrium (LD) across this region. The region of association also contains a relatively gene-rich area including the cytokine gene cluster. Since the identification of this risk haplotype, several groups have independently confirmed the association between this haplotype and CD (Armuzzi et al. 2003; Giallourakis et al. 2003; Mirza et al. 2003; Negoro et al. 2003; Latiano et al. 2006; Torkvist et al. 2007). In some studies association with UC has also been established (Giallourakis et al. 2003; McGovern et al. 2003).

In 2004, two functional polymorphisms 1672C→T and -207G→C within the SLC22A4 (OCTN1) and SLC22A5 (OCTN2) genes, respectively, were proposed to be the critical mutations associating with CD susceptibility in this region (Table 5) (Peltekova et al. 2004). These two polymorphisms were in strong LD, creating a 2-allele risk haplotype (SLC22A-TC) which was enriched among CD patients. Preliminary functional studies of these variants demonstrated altered transcription and transporter activity of various organic cations. After this finding several replication studies in different populations have been performed with conflicting results. Positive association with CD has been obtained in several populations (Table 5) (Gazouli et al. 2005a; Noble et al. 2005a; Torok et al. 2005; Fisher et al. 2006; Leung et al. 2006; Martinez et al. 2006; Onnie et al. 2006a; Torkvist et
al. 2007) and also with IBD in one study (Waller et al. 2006). Many studies have also reported that this association is dependent on the background risk haplotype (Noble et al. 2005a; Fisher et al. 2006; Onnie et al. 2006a; Russell et al. 2006; Waller et al. 2006; Torkvist et al. 2007). Two studies have reported association between SLC22A variants and pediatric CD patients from North America (Babusukumar et al. 2006) and Italy (Cucchiara et al. 2007b) and one with pediatric IBD (Russell et al. 2006). Many studies have resulted in negative association, including for Japanese (Yamazaki et al. 2004; Vermeire et al. 2005) and Flemish (Vermeire et al. 2005) patients.

Although there is controversy over the true causal gene within the IBD5 locus, studies have tried to further assess the specific phenotype associated with this region. As in the case of true variants, findings related to the associated phenotype are inconsistent. Several groups have reported a lack of association between IBD5 and a specific disease location whereas some have reported an association to perianal (Armuzzi et al. 2003; Vermeire et al. 2005), ileocolonic (Gazouli et al. 2005a; Leung et al. 2006), or colonic involvement (Gazouli et al. 2005a). In addition, penetrating (Urcelay et al. 2005; Vermeire et al. 2005; Latiano et al. 2006) and stricturing (Latiano et al. 2006) phenotype associations have been reported with this locus.

Even though the association of the IBD5 locus to CD has been replicated repeatedly, the precise causal variants remain unknown, most likely due to strong LD across this region. Recently, the SLC22A5 gene variant was excluded as the potential causal variant (Silverberg et al. 2007). During the year 2007, GWA studies in CD have also been able to replicate the IBD5 locus and this even further strengthens the association with CD (Consortium 2007; Hampe et al. 2007; Raelson et al. 2007).

5.2 Major histocompatibility complex region (IBD3 locus)

The major histocompatibility complex (MHC) on chromosome 6p21.3 has been the most studied genetic region in IBD (Ahmad et al. 2006). The MHC region, known in humans as the human leukocyte antigen (HLA) region, spans approximately four megabases (Mb) and contains a few hundred genes (Consortium 1999; Yap et al. 2004). The first report of HLA association with IBD was published in 1972 (Gleeson et al. 1972) and since then more than 100 studies have replicated the HLA association in IBD susceptibility. The IBD3 locus on chromosome 6p includes the MHC region and has been linked to both UC and CD susceptibility.
Several independent genome-wide linkage studies have also shown evidence of linkage to \( IBD3 \) (Table 3). The first linkage to chromosome 6 containing the HLA region with the IBD phenotype was reported in 1999 (Hampe et al. 1999a) and soon after linkage to CD was also established (Ma et al. 1999). Since then, several other studies have confirmed the linkage to 6p (Hampe et al. 1999b; Yang et al. 1999b; Rioux et al. 2000; Dechairo et al. 2001; Barmada et al. 2004). This region was further underlined by a meta-analysis of ten published genome scans (van Heel et al. 2004). The study included 1952 IBD affected relative pairs, and \( IBD3 \) was the only locus that gained genome-wide significance. It is also noteworthy that this locus presented stronger evidence of linkage than the \( IBD1 \) locus, which contains the CD susceptibility gene \( CARD15 \). However, more recent GWA studies have not confirmed association with this region. Lack of association is most likely due to the use of CD rather than UC patient material.

The MHC is divided into three different regions: HLA class II (centromeric), class III, and class I (telomeric) (Consortium 1999). Numerous case-control studies in IBD have evaluated association with the classical class I and II alleles because they play a central role in the human immune response. Although many of these associations have been conflicting, association to IBD appear to be with class II rather than class I genes. The classical HLA class II genes encode cell-surface glycoproteins expressed on antigen presenting cells (APC) (Ahmad et al. 2006). The major role of these glycoproteins is to present different peptides (antigens) to T cell receptors (TCR) which in turn results in helper T cell (CD4\(^+\)T) activation. There are three types of HLA class II molecules; HLA-DR, HLA-DQ, and HLA-DP. Each of these molecules is made of \( \alpha \beta \) heterodimers (\( \alpha_1\alpha_2\beta_1\beta_2 \)). The \( \alpha \) chains are encoded by DRA, DQA, and DPQ genes. The first is invariant while the other two are highly polymorphic. The \( \beta \) chains of DRB1, DQB1, and DPB1 are also highly polymorphic. Gene duplication events have occurred in the past producing additional DRB3, 4, and 5 genes and these are expressed at low levels on the cell surface (Yap et al. 2004). Of the HLA class II molecules, HLA-DRB1 is the most extensively studied gene in IBD (Ahmad et al. 2006).

5.2.2 The HLA-DRB1 gene association with IBD

The HLA-DRB1 gene is the most polymorphic of the HLA-DR molecules and is present in all individuals (Stokkers et al. 1999; Annese et al. 2005b). This gene has almost 200 alleles. The most reliable HLA-DRB1 associations were highlighted in a meta-analysis of 29 studies (Stokkers et al. 1999), where significant positive association was achieved in UC with the DRB1*0103 and DRB1*1502 alleles. In CD, positive association was found for the DRB1*0401 and DRB1*0701 alleles.
The HLA-DRB1*0103 allele encodes an HLA-DR molecule with a unique sequence within peptide-binding groove, allowing it to bind and present potentially disease-associated peptides (Coppin et al. 1993). It is a rare allele with less than 2% frequency in Caucasian populations but still this is the most replicated HLA class II allele in UC susceptibility (Ahmad et al. 2006). It is also present at similar frequency in the Jewish population but absent from the Japanese, the same phenomenon observed with CARD15 mutation frequencies. A strong association between this rare allele and UC was introduced for the first time in 1996 (Satsangi et al. 1996b), and a few years later also with CD (Trachtenberg et al. 2000). In the meta-analysis, three substudies also showed a moderate association in unselected UC patients (Stokkers et al. 1999). The association of this rare allele with UC has been later confirmed in different populations, namely in Dutch (Bouma et al. 1999), Spanish (de la Concha et al. 1999), North American (Trachtenberg et al. 2000; Silverberg et al. 2003), British (Ahmad et al. 2003), and Mexican (Yamamoto-Furusho et al. 2003) cohorts. In the first study, association with the rare allele was strongest among male patients with extensive disease (Satsangi et al. 1996b). This allele was also associated with severe disease requiring colectomy. In another study, this same allele was found more frequently in UC patients who underwent the surgery (Roussomoustakaki et al. 1997). Extensive disease involvement and extraintestinal manifestations were also more common among DRB1*0103 allele carriers. Other studies have further confirmed both extensive colitis and severe disease as defined by a need for (procto)colectomy among DRB1*0103 positive UC patients (Bouma et al. 1999; Ahmad et al. 2003; Yamamoto-Furusho et al. 2003). This uncommon allele may also be associated with a shorter mean time for surgery among patients who require colectomy (Ahmad et al. 2003).

CD is also associated with the DRB1*0103 allele in North American (Trachtenberg et al. 2000; Silverberg et al. 2003; Newman et al. 2004) and Spanish populations (Fernandez et al. 2004). The frequency of the DRB1*0103 allele was 10 fold higher in CD patients with colonic involvement than in healthy controls (Silverberg et al. 2003). This association has been replicated later in other studies (Fernandez et al. 2004; Newman et al. 2004). Perianal involvement and penetrating behaviour have also been shown to associate with this uncommon allele in two studies (Ahmad et al. 2002; Fernandez et al. 2004). It is quite clear now that this rare DRB1*0103 allele is a risk factor for both UC and colonic CD, suggesting it may have a role in chronic inflammation of the colon which is independent of UC or CD.
AIMS OF THE STUDY

The aims of the present study were to search for genes and their alterations predisposing to Crohn’s disease and ulcerative colitis, the two major forms of inflammatory bowel diseases, to define genotype-phenotype relationships in IBD, and to delineate functional consequences of the genetic variations detected.

The specific aims were to

1. evaluate the allele frequencies of the common $CARD15$ mutations in Finnish IBD patients, and search for possible associations between genotype and phenotype.
2. search for novel $CARD15$ mutations in Finnish CD patients, to assess their functional consequences, and to relate their presence with the clinical phenotype.
3. participate into an international collaboration aimed at evaluating the allele frequencies and disease-related risks of the common $CARD15$ mutations in the population.
4. identify variants of other candidate genes that are associated with risk or phenotypes of IBD.
PATIENTS AND METHODS

1. STUDY SUBJECTS

1.1 Patients with inflammatory bowel disease

Consecutive IBD patients treated at the departments of Gastroenterology and Surgery of the Helsinki University Hospital and Maria Hospital (a municipal hospital of Helsinki) during the years 1995-1998 or consecutive UC patients who had undergone colectomy in the 1990s at the Helsinki University Hospital were asked to participate in the study (Halme et al. 2002). Familial patients were also recruited from other parts of the country with the assistance of colleagues from the other University Hospitals in Finland. The diagnosis of CD and UC was based on standard endoscopical and histological criteria (Lennard-Jones 1989). A questionnaire on the history and details of their disease as well as the number of their affected and non-affected relatives were given to 1000 patients; 82% of the patients returned the questionnaire and gave their consent to the study. Patients were considered as familial IBD cases if they had at least one affected first-degree family member. In families with two or more affected members, the first member returning the questionnaire was termed the proband. Patients reporting parents originating outside Finland and affected relatives more distant than first-degree relatives were excluded. The final material of the present study consisted of 754 IBD patients, 459 of them being UC and 295 CD patients (Table 7).

Table 7. Number of patients and healthy controls used in studies I-V.

<table>
<thead>
<tr>
<th>Study</th>
<th>CD</th>
<th>UC</th>
<th>Mixed families</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Familial</td>
<td>Sporadic</td>
<td>Familial</td>
<td>Sporadic</td>
</tr>
<tr>
<td>I</td>
<td>46</td>
<td>198</td>
<td>37</td>
<td>62</td>
</tr>
<tr>
<td>II</td>
<td>44</td>
<td>196</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>III</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IV</td>
<td>initial cohort</td>
<td>20</td>
<td>-</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>extended cohort</td>
<td>44</td>
<td>195</td>
<td>78</td>
</tr>
<tr>
<td>V</td>
<td>44</td>
<td>196</td>
<td>115</td>
<td>344</td>
</tr>
</tbody>
</table>

*190 blood donors and 6 healthy individuals from laboratory staff

Details on the course of the disease were collected from the patient records. The location of inflammation was defined according to the largest macro- and microscopic extent registered at some time during the history of the disease. Location of inflammation in UC was classified as either left-sided or extended. For classifying the severity of the disease,
the need for medication and surgery was taken into account. CD was graded according to the Vienna classification (Gasche et al. 2000).

1.2 Healthy controls

DNA samples from anonymous healthy blood donors served as controls for the estimation of the allele frequencies in the background population. In studies I-V, varying numbers of blood donors were used as controls (Table 7). The samples were kindly provided by Dr. Tom Krusius, the Finnish Red Cross Blood Transfusion Service, Helsinki, Finland. Blood donors were from the capital city area of Helsinki. Mean age of the blood donors was 45 years at blood sampling, with equal numbers of women and men.

In study III, 3575 healthy individuals of Caucasian origin recruited by 15 groups from three different continents were studied for the three common \textit{CARD15} mutations. Healthy individuals were a mixture of blood donors, healthy spouses of affected patients, and control groups from other studies. Only information on their European ancestry and apparent health was available for all cohorts. The number of healthy controls in different cohorts varied between 45 and 587. The Finnish cohort consisted of 300 blood donors. For functional \textit{in vitro} studies in study II, six volunteers from the laboratory staff were used as healthy controls.

1.3 Ethical aspects

Informed written consent was obtained from all study subjects. In addition, the local Ethics Review Committee of the Department of Medicine, Helsinki University Hospital, had approved these studies.
2. METHODS

The methods used in the present study are described in more detail in the original publications I to V. They are also combined into Table 8.

Table 8. Methods used in the original publications in this thesis. The numbers of the original publications are indicated by their Roman numerals I-V.

<table>
<thead>
<tr>
<th>Method</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA extraction</td>
<td>I, II, III, IV, V</td>
</tr>
<tr>
<td>PCR</td>
<td>I, II, III, IV, V</td>
</tr>
<tr>
<td>RFLP</td>
<td>I, II, III, V</td>
</tr>
<tr>
<td>dHPLC</td>
<td>II, IV</td>
</tr>
<tr>
<td>Sequencing/Pyrosequencing</td>
<td>II, III, IV</td>
</tr>
<tr>
<td>Real time quantitative PCR</td>
<td>III</td>
</tr>
<tr>
<td>MassARRAY by Sequenom</td>
<td>III, V</td>
</tr>
<tr>
<td>Recombinant DNA techniques (cloning)</td>
<td>IV</td>
</tr>
<tr>
<td>Functional studies <em>in vitro</em></td>
<td>II</td>
</tr>
<tr>
<td>ELISA</td>
<td>II</td>
</tr>
<tr>
<td>Bioinformatic analysis</td>
<td>I, II, IV, V</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>I, II, III, IV, V</td>
</tr>
</tbody>
</table>
1. MUTATIONS IN THE CARD15 GENE AND THEIR RELATION TO CLINICAL PHENOTYPES

1.1 The CARD15 gene mutation spectrum (I, II)

When this study was initiated, the association of the CARD15 gene with CD was already established. Three mutations, R702W, G908R, and 1007fs, in the CARD15 gene responsible for CD susceptibility were investigated in the Finnish cohort of 271 CD patients, 99 UC patients, and 300 blood donors as background population. CD patients were from three different categories based on their family relationships. Forty-six of these patients were familial cases and 198 were without family history of CD. The remaining 27 CD patients were from mixed IBD families. Results showed that 15.5% of CD patients, 9.1% of UC patients, and 6.7% of controls carried at least one of the three CARD15 mutations. Allele frequencies of the common mutations are summarised in Table 9. The allele frequency of the 1007fs mutation was significantly higher in CD patients than in healthy blood donors (4.8% vs. 1.7%, p<0.01) and the corresponding frequency in familial probands was also statistically different when compared to sporadic patients (10.9% vs. 3.5%, p<0.01). The combined frequency of the three mutations between CD patients and controls was also significantly different (8.7% vs. 3.5%, p<0.01). One sporadic and two familial CD patients were homozygous for the 1007fs mutation and their affected siblings were homozygotes as well. Three compound heterozygotes were also identified, two of them being CD patients with the R702W and G908R mutations and one control carrying the R702W and 1007fs mutations.

### Table 9. Allele frequencies of the three common CARD15 mutations in patients with CD and UC and in blood donors.

<table>
<thead>
<tr>
<th>Group</th>
<th>R702W</th>
<th>G908R</th>
<th>1007fs</th>
<th>Together</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD sporadic (n=198)</td>
<td>0.033</td>
<td>0.005</td>
<td>0.035</td>
<td>0.073</td>
</tr>
<tr>
<td>CD familial (n=46)</td>
<td>0.033</td>
<td>0.011</td>
<td>0.109</td>
<td>0.153</td>
</tr>
<tr>
<td>CD mixed families (n=27)</td>
<td>0.037</td>
<td>0.000</td>
<td>0.037</td>
<td>0.074</td>
</tr>
<tr>
<td>CD all (n=271)</td>
<td>0.033</td>
<td>0.006</td>
<td>0.048</td>
<td>0.087</td>
</tr>
<tr>
<td>UC (n=99)</td>
<td>0.015</td>
<td>0.000</td>
<td>0.030</td>
<td>0.045</td>
</tr>
<tr>
<td>Blood donors (n=300)</td>
<td>0.018</td>
<td>0.000</td>
<td>0.017</td>
<td>0.035</td>
</tr>
</tbody>
</table>

In the subsequent study (study II), a total of 240 unrelated patients with CD were screened for additional mutations in the CARD15 gene by dHPLC and direct sequencing. A total of 30 different sequence variants were identified of which 25 were in the exonic and 5 in the

---

52
untranslated or intronic regions (Table 10). Eighteen of 25 exonic variants were predicted to result in amino acid changes. Thirteen of these amino acid substitutions, including the three common mutations, were previously described, while five were novel.

Of the sequence variants occurring in more than a single individual, only the frequency of the 1007fs mutation was significantly higher in patients (4.8%) than in controls (1.6%, p=0.01) in study II. The allele frequencies of the R702W and G908R mutations in CD patients were similar to those in study I (3.3% and 0.4%, respectively) and the corresponding frequencies in controls were 1.6% and 0%, respectively. There was no significant association of these two mutations with the risk of CD in accordance with previous results (study I).

The five novel amino acid substitutions R38M, W355X, P727L, W907R, and R1019X each occurred in a sporadic CD patient but in none of the healthy blood donors (Figure 2). Two of these mutations were nonsense (W355X and R1019X) and the remaining three were missense mutations (R38M, P727L, and W907R). The W355X mutation co-occurred with the 1007fs mutation, while the R1019X mutation occurred alone. The patient with the R38M mutation was also a homozygote for the 1007fs mutation, indicating that the R38M mutation had once occurred in the allele containing the 1007fs alteration. Furthermore, the patient carrying the P727L variant was also carrying the 1007fs mutation; again, these two mutations were on the same chromosome as confirmed by DNA analysis of the patient’s daughter. The third patient with the W907R variant was also heterozygous for the T294S variant.

The previously described amino acid changes T189M, N289S, R791Q, and M863V were each found in one individual patient but none of the controls. It is again of note that the patient with N289S as well as the patient with M863V was a carrier of the common 1007fs mutation, but the allelic relation of these variants could not be determined. The remaining five amino acid substitutions were absent in CD patients (A660G) or occurred in similar frequencies in patient and control groups (T294S, R703C, A755V, and V955I), suggesting that they may not be related to CD susceptibility.

Seven nucleotide changes (C534G, A621T, C1377T, T1761G, C1833T, G2406T, and C2862T) which did not change amino acids were also found (Table 10). Their allele frequencies did not differ between patients and controls, and thus these variations were considered as silent polymorphisms. Five noncoding sequence variants were also identified, four of them in the intervening sequences (IVS2-29T→G, IVS2-25G→T,
IVS4+10A→C, and IVS7-15delCT) and one in the 5’ untranslated region (5’UTR-59G→A). There was no difference in their frequencies between patients and controls.

**Table 10.** *CARD15* mutations and polymorphisms identified in patients with CD and the background population. Novel amino acid changes are shown in bold. CARD=caspase recruitment domain, NBD=nucleotide-binding domain, LRR=leucine-rich repeat.

<table>
<thead>
<tr>
<th>Nucleotide change</th>
<th>Amino acid change</th>
<th>Protein domain</th>
<th>CD patients (n=250)</th>
<th>Healthy controls (n=190)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-59G&gt;A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-29T&gt;G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-25G&gt;T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G113T</td>
<td>R38M</td>
<td>CARD1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C534G</td>
<td>S178S</td>
<td>CARD2</td>
<td>195</td>
<td>151</td>
</tr>
<tr>
<td>C566T</td>
<td>T189M</td>
<td>CARD2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A621T</td>
<td>P207P</td>
<td>CARD2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C802T</td>
<td>P268S</td>
<td></td>
<td>98</td>
<td>63</td>
</tr>
<tr>
<td>A866G</td>
<td>N289S</td>
<td>NBD</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C881G</td>
<td>T294S</td>
<td>NBD</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>G1065A</td>
<td>W355X</td>
<td>NBD</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C1377T</td>
<td>R459R</td>
<td>NBD</td>
<td>104</td>
<td>63</td>
</tr>
<tr>
<td>T1761G</td>
<td>R587R</td>
<td></td>
<td>185</td>
<td>145</td>
</tr>
<tr>
<td>C1833T</td>
<td>A611A</td>
<td></td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>C1979G</td>
<td>A660G</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C2104T</td>
<td>R702W</td>
<td></td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>C2107T</td>
<td>R703C</td>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>C2180T</td>
<td>P727L</td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C2264T</td>
<td>A755V</td>
<td>LRR1</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>G2372A</td>
<td>R791Q</td>
<td>LRR2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>G2406T</td>
<td>V802V</td>
<td>LRR2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>IVS4+10A&gt;C</td>
<td></td>
<td></td>
<td>26</td>
<td>20</td>
</tr>
<tr>
<td>A2587G</td>
<td>M863V</td>
<td>LRR5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>IVS7-15delCT</td>
<td></td>
<td></td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>T2719C</td>
<td>W907R</td>
<td>LRR6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>G2722C</td>
<td>G908R</td>
<td>LRR6</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>C2862T</td>
<td>N954N</td>
<td>LRR8</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>G2863A</td>
<td>V955I</td>
<td>LRR8</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>3020insC</td>
<td></td>
<td>LRR10</td>
<td>23</td>
<td>6</td>
</tr>
<tr>
<td>C3055T</td>
<td>R1019X</td>
<td>LRR10</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
1.2 The CARD15 mutation frequency in the background population (III)

The allele and genotype frequencies were estimated for the R702W, G908R, and 1007fs mutations in 3575 healthy people of Caucasian origin. This background population was obtained from 15 different groups in Europe, North America, and Australia. The combined allele frequencies of the R702W, G908R, and 1007fs mutations were 4.3%, 1.2%, and 2.3%, respectively (Table 11), with large geographic fluctuations of the G908R, 1007fs, and wild type alleles (p<0.0001). Comparison of individual cohorts showed that the G908R mutation was absent in Finland and also very rare in Scotland but frequent in Italy and California (Table 11). The 1007fs mutation was in contrast rare in Italy and Brisbane but it was frequently found in Cambridge as well as in Germany and Chicago (Table 11). Comparison of wild type allele frequencies showed that CARD15 mutations were rare in Finland and Brisbane, while they were common in Belgium and Cambridge (Table 11).

Among 3575 healthy controls, 18 were double mutation carriers. According to Hardy-Weinberg equilibrium, the expected number in this cohort was 23 but this deviation was not statistically significant. This provided strong evidence that the penetrance of the common mutations is low. Finally, in a case of Mendelian inheritance with limited heterogeneity, the disease incidence is expected to be directly related to mutation frequency. Thus, the relationship between the annual incidence and the CARD15 mutation frequency in different populations were explored. No simple relationship was observed between these parameters for the cohorts studied.
Table 11. Minor allele frequencies of the three common CARD15 mutations in apparently healthy individuals across three continents.

<table>
<thead>
<tr>
<th>Population</th>
<th>Controls (n)</th>
<th>R702W</th>
<th>G908R</th>
<th>1007fs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td>300</td>
<td>0.018</td>
<td>0.000</td>
<td>0.017</td>
<td>0.035</td>
</tr>
<tr>
<td>Scotland</td>
<td>136</td>
<td>0.048</td>
<td>0.004</td>
<td>0.018</td>
<td>0.070</td>
</tr>
<tr>
<td>England (Oxford)</td>
<td>587</td>
<td>0.042</td>
<td>0.011</td>
<td>0.019</td>
<td>0.072</td>
</tr>
<tr>
<td>England (Cambridge)</td>
<td>351</td>
<td>0.057</td>
<td>0.010</td>
<td>0.041</td>
<td>0.108</td>
</tr>
<tr>
<td>Belgium</td>
<td>144</td>
<td>0.066</td>
<td>0.021</td>
<td>0.028</td>
<td>0.115</td>
</tr>
<tr>
<td>Germany</td>
<td>370</td>
<td>0.045</td>
<td>0.007</td>
<td>0.038</td>
<td>0.090</td>
</tr>
<tr>
<td>Italy</td>
<td>205</td>
<td>0.041</td>
<td>0.027</td>
<td>0.007</td>
<td>0.075</td>
</tr>
<tr>
<td>France</td>
<td>174</td>
<td>0.049</td>
<td>0.011</td>
<td>0.026</td>
<td>0.086</td>
</tr>
<tr>
<td>Canada (Quebec)</td>
<td>58</td>
<td>0.052</td>
<td>0.009</td>
<td>0.009</td>
<td>0.070</td>
</tr>
<tr>
<td>Canada (Toronto)</td>
<td>100</td>
<td>0.050</td>
<td>0.020</td>
<td>0.030</td>
<td>0.100</td>
</tr>
<tr>
<td>USA (Chicago)</td>
<td>273</td>
<td>0.037</td>
<td>0.016</td>
<td>0.038</td>
<td>0.091</td>
</tr>
<tr>
<td>USA (California)</td>
<td>220</td>
<td>0.036</td>
<td>0.030</td>
<td>0.016</td>
<td>0.082</td>
</tr>
<tr>
<td>USA (Mid-Atlantic)</td>
<td>45</td>
<td>0.033</td>
<td>0.011</td>
<td>0.033</td>
<td>0.077</td>
</tr>
<tr>
<td>Australia (Brisbane)</td>
<td>203</td>
<td>0.039</td>
<td>0.007</td>
<td>0.007</td>
<td>0.053</td>
</tr>
<tr>
<td>Australia (Canberra)</td>
<td>409</td>
<td>0.045</td>
<td>0.007</td>
<td>0.010</td>
<td>0.062</td>
</tr>
<tr>
<td>Total</td>
<td>3575</td>
<td>0.043</td>
<td>0.012</td>
<td>0.023</td>
<td>0.078</td>
</tr>
</tbody>
</table>

1.3 Phenotypic characteristics of CARD15 mutations (I, II)

The phenotypic characteristics of the CD patients were assessed according to the Vienna classification (Table 2) (Gasche et al. 2000). In addition, patient gender, familial status, duration of the disease, medication, indication for surgery, and smoking behaviour were investigated. Upon analysis of the three common mutations, CD patients with one or two mutations more often had ileal disease location (90% vs. 73%, p<0.05). In addition, these patients also more often had a complicated form of the disease as indicated by stricturing and penetrating behaviour (88% vs. 56%, p<0.01). Median disease duration was also two times longer among mutation carriers than non-carriers (12.3 vs. 6.0 years, p<0.01); however, upon comparison of ages at diagnoses no significant difference was seen. The need for bowel surgery was more common among mutation carriers than among non-carriers (67% vs. 46%, p<0.05).

Two familial CD patients and their affected siblings were homozygous for the 1007fs mutation. These four CD patients shared some phenotypic characteristics, including young age at diagnosis (≤20 years), an ileal location, and stricturing behaviour of the disease. One single sporadic CD patient who was homozygous for the 1007fs mutation also had the right colon affected in addition to the ileum. In this case, the diagnosis was established at the age of 13, and strictures had complicated the disease. He had also undergone surgical interventions three times.
In the CARD15 mutation screening study (study II), all five CD patients with novel CARD15 mutations had a complicated form of ileal or ileocolonic disease. In addition, R38M and W907R mutation carriers had short bowel syndrome as a complication of several bowel resections. It should also be noted that four of these five patients carried at least one additional CARD15 mutation.

Since the number of mutation carriers was higher in study II than in study I, genotype-phenotype associations were reanalysed for possible new correlations. A variant was considered a mutation if it resulted in a truncated protein or changed an amino acid that was conserved between species or if the present or previous functional studies had demonstrated impaired function for this protein. Reanalysis did not change the previous results concerning the association with ileal involvement (p=0.03). The complicated form of the disease was associated with double mutation carriers (p=0.045), further strengthening the previous results. Based on the fact that the 1007fs mutation was the only variant significantly associated with CD, it was important to study how this mutation affects phenotypic associations. Removal of the 1007fs positive patients retained the negative association with pure colonic disease (p=0.011) but the association with ileal location was no longer significant (p=0.11). However, the complicated form of the disease was still associated with mutation carriers as compared to non-carriers (p=0.036).

2. FUNCTIONAL STUDIES OF THE CARD15 GENE (II)

In order to characterise the functional consequences of the four novel mutations R38M, W355X, P727L, and R1019X in the CARD15 gene in vitro, mononuclear cells of patients were cultured with and without MDP, a ligand for NOD2, and production of interleukin 8 (IL-8) was determined. The fifth novel mutation W907R was not studied because the patient had uraemia. In addition, six wild type CD patients, another six 1007fs heterozygous patients, one compound heterozygote for the R702W and 1007fs mutation, and two homozygotes for the 1007fs mutation were included in the functional studies. Six individuals, negative for the three CARD15 mutations, from the laboratory staff were used as healthy controls in this functional study. The immune inflammatory status of the individuals was also determined: no significant difference in inflammation status between patient and control groups was observed (study II, Table 2). In addition, no significant difference was observed in lymphocyte and monocyte counts or the proportion of CD14-positive cells between patients and controls.

The results of IL-8 responses showed that three groups, i.e. wild type patients, 1007fs heterozygous patients, and healthy controls as well as a compound heterozygote, increased
Figure 3. Results of functional studies of known CARD15 mutations are shown as interleukin 8 responses of peripheral blood mononuclear cells stimulated by muramyl dipeptide (MDP). Data show median and range (bars) of six subjects and the symbols without bars indicate individual patients.

Figure 4. Results of functional studies of novel CARD15 mutations are shown as interleukin 8 responses of peripheral blood mononuclear cells stimulated by muramyl dipeptide (MDP). The healthy control curve is shown again to permit orientation. All novel CARD15 mutation curves represent individual patients.
IL-8 production along with MDP concentration in a comparable manner (Figure 3). The corresponding IL-8 levels of each individual are also represented as area under a curve (AUC) using the trapezoidal rule (Figure 5), with no significant difference in AUC values between these three groups. On the contrary, the two 1007/fs homozygotes had consistently low IL-8 levels (Figure 3), and the respective % AUC values (patient AUC value as a percentage of the lowest control AUC value) were decreased to 12.1% and 5.4% (Figure 5). The range of % AUC values of the 1007/fs heterozygotes was over 100%, indicating that 1007/fs heterozygosity was not sufficient to reduce IL-8 production.

![Area under a curve (AUC) values of functional studies.](image)

Figure 5. Area under a curve (AUC) values of functional studies. AUC value stands for the serial measurements of IL-8 levels (pg/ml) in response to different muramyl dipeptide (MDP) concentrations (0-1000 ng/ml). Each column symbolizes a single individual. Novel amino acid variations are; A denotes R38M+1007/fs/1007/fs, B wt/P727L+1007/fs, C W355X/1007/fs, and D wt/R1019X.

Of novel mutations, the patient carrying the W355X variant in conjunction with the 1007/fs mutation had low IL-8 responses similar to two 1007/fs homozygotes (Figure 4). This indicates that the two mutations are on different chromosomes and both of these mutations contribute to impaired IL-8 response. The % AUC value of this patient was only 7.0% (Figure 5). The patient with the R38M mutation also had low IL-8 responses (Figure 4) with a 10.1% AUC value (Figure 5) but a firm conclusion about the effect of this mutation can not be made because he was also homozygous for the 1007/fs mutation. Finally, the P727L and R1019X mutations were also associated with depressed IL-8 production (Figure 4) with % AUC values of 61.4 and 61.9, respectively (Figure 5). It is also of note that the P727L patient was carrying the 1007/fs mutation on the same allele.
3. GENETIC VARIATIONS IN CANDIDATE GENES AND CORRELATION TO PHENOTYPE

3.1 Screening of the TRAF6 gene (IV)

Since TRAF6 presented as a potential IBD candidate gene based on both functional data and the genomic location of the gene in the chromosome 11 IBD linkage region (Paavola-Sakki et al. 2003), 95 unrelated IBD patients were screened for mutations in this gene by dHPLC. Only one intronic variation was found, a single nucleotide insertion/deletion polymorphism containing either eight or seven thymine bases (8T/7T), respectively. The location of this variation was eight bases from the intron 3/exon 4 boundary, maintaining the possibility that it affects mRNA splicing. 290 additional CD patients and 416 UC patients were then genotyped for this variation, along with 320 blood donors to obtain a background frequency in the population. Allele and genotype frequencies between different groups were not significantly different. In addition, several clinical characteristics of the patients were analysed because this polymorphism could theoretically affect a sub-phenotype of the disease without affecting the risk of getting the disease itself. Clinical characteristics analysed included the same parameters as in study I. No significant relationship between the polymorphism and IBD phenotypes were found.

3.2 The rare HLA-DRB1*0103 allele (V)

Previous studies have shown strong association of the rare HLA-DRB1*0103 allele with UC, and later also with colonic CD (Ahmad et al. 2006). In the present study, 15 UC (3.3%, n=459) and three CD (1.3%, n=240) patients carried the HLA-DRB1*0103 allele in a heterozygous form. In contrast, none of the 190 blood donors carried this allele. The allele frequencies were statistically different between UC and controls (p=0.008) as well as between IBD and controls (p=0.019), but not between CD patients and controls (p=0.26) (study V, Table 3). HLA-DRB1*0103 positive UC patients tended to more often have extensive colitis (67% vs. 61%), and they also needed colectomy more often (47% vs. 29%), but the disease duration was also longer (median years 13.8 vs. 9.7) when compared to patients without this rare allele. These phenotypic differences, however, were not statistically different. All three HLA-DRB1*0103 positive CD patients had a colonic location of the disorder.

3.3 Eight SNPs spanning the IL23R gene (V)

Recently, a genome-wide association study reported a highly significant association between CD and IL23R (Duerr et al. 2006). In the present study, a coding variant R381Q
and seven additional noncoding SNPs in the *IL23R* gene and the intergenic region between the *IL23R* and *IL12RB2* genes were studied in 699 IBD patients and 292 healthy blood donors. Two additional markers rs7517847 and rs1495965 from the original study were excluded due to technical problems. Five SNPs, including the coding variant, were significantly associated with CD (n=240) (study V, Table 1). The strongest association was observed with marker rs2201841 (p=0.002), where the frequency of the C allele was higher in CD patients than in healthy blood donors (37.2% vs. 28.8%, respectively). The other two disease-associated markers rs10889677 and rs1004819, which were in strong LD with rs2201841, had p values of 0.003 and 0.005, respectively (study V, Table 1). The coding variant, rs11209026, was only marginally associated with CD (p=0.045) and the fifth disease-associated marker rs11465804 had a p value of 0.016. After permutation testing the three strongest markers still showed significant association. The results also showed significant association of three markers rs10889677, rs2201841, and rs1004819 with IBD (p=0.026, p=0.029, and p=0.037, respectively) (study V, Table 1).

Haplotype analysis showed nine distinct haplotypes, two of them associating with CD (study V, Table 2). The most common haplotype (TTCTGCAA) was observed to be a risk haplotype with a frequency of 28.8% in CD patients and 23.2% in blood donors (p=0.040). The rare haplotype (CCTGATCG) tagged with the coding variant rs11209026-A allele, which has been shown to confer strong protection against CD, had frequencies of 0.4% in patients and 2.5% in controls (p=0.008).

Genotype-phenotype correlations showed that three *IL23R* markers (rs2201841, rs10889677, and rs1004819) were also associated with longer disease duration among CD patients. Duration with rs2201841 was 5.7 years in wild types, 7.0 in heterozygotes, and 8.1 in homozygotes (p=0.011). The corresponding years with rs10889677 were 5.8, 6.9, and 11.2 (p=0.013) and with rs1004819 5.8, 7.0, and 12.1, respectively (p=0.039). An association between CD and the stricturing phenotype was also observed with the markers rs10889677 and rs2201841. Thus, CD patients with the minor alleles A of rs10889677 and C of rs2201841 more often had strictureting behaviour than patients without these alleles (p=0.010 and p=0.017, respectively).

3.4 Three SNPs within the *TNFRSF1A* gene (V)

Another positional and functional candidate gene *TNFRSF1A* on chromosome 12p13 was analysed by genotyping previously reported polymorphisms: a 5'UTR -609G→T, a coding 36A→G (P12P), and an intronic IVS6+10A→G variant. The allele and genotype frequencies of these three variants did not differ between patient and control groups (study
RESULTS

V, Table 3). However, when the familial (n=114) and sporadic (n=342) UC patients were analysed separately, an association (p=0.03) between familial UC and the rare coding variant 36G was detected. The extension of healthy control material from 190 to 312 individuals further strengthened this association (p=0.007). The association of the other two polymorphisms with familial UC patients, however, did not reach statistical significance.

In genotype-phenotype analysis, the genotype frequencies of the functional variant A36G also differed significantly (p=0.001) between sporadic and familial UC. Accordingly, the AA (wild type) genotype was more frequent in sporadic cases (31.6% vs. 20.2%) whereas the GG (homozygote) genotype was more common in familial cases (32.5% vs. 17.3%). The intronic variant IVS6+10A→G was also associated with familial UC (p=0.042) with AA (wild type) frequencies of 27.2% in familial and 36.0% in sporadic UC. The GG (homozygote) genotype was more common in familial cases (25.4% vs. 15.8%).

Comparison of other phenotypic characteristics showed that CD patients carrying the rare coding 36G allele as well as the intronic allele IVS6+10G more often had ileocolonic location than did patients without these two variants (p=0.021 and p=0.028, respectively).

3.5 SNPs in genes ATG16L1, DLG5, ABCB1, TLR4, and the chromosome 5 risk haplotype containing SLC22A4 and SLC22A5 (V)

A selection of SNPs in the ATG16L1, DLG5, ABCB1/MDR1, TLR4, and chromosome risk haplotype containing genes SLC22A4 and SLC22A5 were studied for IBD susceptibility. One or two SNPs per candidate gene were studied in a cohort of 699 IBD patients and 190 healthy blood donors. None of these SNPs showed any significant association in the case-control analysis (study V, Table 3). Likewise, carriage of the SLC22A4/5 TC risk haplotype, which has been shown to associate with CD (Peltekova et al. 2004), did not differ between patient and control groups.

Genotype-phenotype comparisons showed that the TLR4 gene polymorphism 299G carriers (72.7%) had more often extensive UC than did D299 carriers (59.6%) (p=0.042).
DISCUSSION

Crohn’s disease and ulcerative colitis, collectively known as IBD, are mostly diseases of the twentieth century. Even though the causes of these two diseases are still incompletely understood, the current model is that the intestinal flora drives an aberrant intestinal immune response and inflammation in the genetically susceptible host (Sands 2007). Interest in the genetic of IBD has grown progressively since the familial occurrence of IBD was brought to our attention in the 1960s (Ahmad et al. 2001). Molecular genetic studies of IBD started as candidate gene studies already in the 1980s (Orchard et al. 2000) but not until a decade later when linkage studies became feasible was real progress in the genetics of IBD achieved. As a result of linkage studies the first CD susceptibility gene, CARD15, was found in 2001. The latest progress in molecular genetics has been the advent of GWA studies which have generated enormous data particularly for IBD in a short period of time.

1. GENETIC VARIATIONS IN THE CARD15 GENE

Three common CARD15 mutations have been shown to associate with CD in Caucasians. Unlike in previous studies, only 15.5% of Finnish CD patients carried any of the three common mutations. This figure was significantly lower than the corresponding estimates for populations from North America or central and southern Europe. However, the present results are similar to those reported from other Scandinavian countries and Scotland (Arnott et al. 2004; Vind et al. 2005; Torkvist et al. 2006; Ernst et al. 2007), which is compatible with our roots (Norio 2003). The present data further confirm the association between the 1007fs mutation and CD but there were no significant differences in the allele frequencies of the R702W and G908R mutations between CD patients and the background population. In Sweden and Denmark, two of three common mutations were independently associated with CD (Vind et al. 2005; Torkvist et al. 2006). In Norway, the R702W and G908R variants had the same allele frequencies among cases and controls, while the 1007fs variant had higher mutation frequency in patients but the difference was not statistically significant (Medici et al. 2006). In conclusion, the present study on Finnish CD patients emphasizes the importance of the 1007fs mutation as a risk allele conferring susceptibility to CD.

In the present study, the 1007fs mutation frequency was higher in familial than sporadic patients. It is also of note that two familial probands out of 46 patients and one patient among 198 sporadic patients were homozygous for the 1007fs mutation. Higher frequencies of the common CARD15 mutations among familial patients has also been
reported for other populations (Cuthbert et al. 2002; Zhou et al. 2002; Oostenbrug et al. 2006b) but some studies have been unable to confirm this (Newman et al. 2004; Annese et al. 2005a; van der Linde et al. 2007).

In the present CARD15 screening study, 30 different sequence variants were detected, half the number of variants found in a large European study (Lesage et al. 2002). There was no evidence for a putative Finnish founder or major mutation, but instead five novel mutations (R38M, W355X, P727L, W907R, and R1019X) with apparent pathophysiological roles in CD were detected. The 1007fs variant remained the only mutation that associated significantly with CD. All novel mutations except the W907R variant were tested for functional consequences in vitro.

Two novel variants (W355X and R1019X) were nonsense mutations like the common 1007fs mutation. Functional studies on the W355X mutation showed that this mutation has an inhibitory effect on IL-8 production similar to the 1007fs mutation. It was impossible to judge whether the W355X and 1007fs mutations were on the same chromosome since other family members were unavailable. Nevertheless, the poor IL-8 production similar to 1007fs homozygotes strongly suggests that these two mutations are on different alleles. R1019X is very similar to the 1007fs mutation in that they both reside in the LRR domain. The patient carrying the R1019X mutation showed lower IL-8 production than did the 1007fs heterozygous patients, indicating that the R1019X mutation could be the more severe of these two mutations. It is interesting that van Heel and co-workers found one CD patient carrying the R1019G mutation (van Heel et al. 2006). This mutation, however, was not shown to have impaired IL-8 production.

The other two novel variants R38M and P727L were missense mutations. Since the R38M mutation occurred in a 1007fs homozygous patient, firm conclusions about its functional consequences can not be made. The P727L mutation also co-occurred with the 1007fs mutation. However, high degrees of cross-species conservation of the respective wild type amino acids at these positions and the chemical nature of the changes involved suggest the pathogenic significance of these variants. Substitution of arginine to methionine at position 38 replaces a hydrophilic and polar amino acid to a neutral and nonpolar one. Replacement of proline by leucine at position 727 may radically change the overall structure of the encoded protein. Furthermore, functional studies showed that the patient with mutations P727L and 1007fs on the same chromosome had, unlike the pure 1007fs heterozygotes, strongly depressed IL-8 production. This was most likely due to presence of two mutations on the same chromosome.
The fifth novel variant W907R most likely has impaired function based on the previous observation made for an analogous mutation W907L. This particular mutation has shown to be a loss of function mutation as W907 is involved in MDP binding and/or protein-protein interaction required for MDP recognition (Tanabe et al. 2004). In addition, W907 is a conserved amino acid based on cross-species comparisons.

There are some functional data available for the previously reported variants N289S, T294S, R703C, M863V, and V955I (Lesage et al. 2002; Miceli-Richard et al. 2002; Sugimura et al. 2003a; Sun et al. 2003; Lakatos et al. 2004; Tukel et al. 2004; Lakatos et al. 2005; King et al. 2006; Schnitzler et al. 2006; van Heel et al. 2006), identified also in the present study, as well as for the common mutations R702W and G908R. Human embryonic kidney (HEK) 293 cells transfected with R702W and G908R were characterised by at least a two fold reduction in both basal NF-κB activity and PGN-induced response, indicating a constitutive defect in CARD15 function (Bonen et al. 2003; Chamaillard et al. 2003). The rare variant N289S which was present in one CD patient has been shown to have similar functional characteristics (Chamaillard et al. 2003). On the contrary, the responses of T294S, R703C, M863V, and V955I variants to PGN were similar to the wild type allele (Chamaillard et al. 2003). Additional data supporting the normal function of the R703C variant comes from mononuclear cell study which showed no profound loss of MDP-sensing with this variant (van Heel et al. 2006). Some studies, however, have suggested that the R703C variant is a potential disease-susceptibility allele based on sequence alignments, use of different prediction programs, and case-control analysis (Lesage et al. 2002; Sun et al. 2003; Tukel et al. 2004; Lakatos et al. 2005; King et al. 2006).

The previously described variants A755V (Lesage et al. 2002; Miceli-Richard et al. 2002; Lakatos et al. 2005; Schnitzler et al. 2006; van Heel et al. 2006) and R791Q (Miceli-Richard et al. 2002; Sugimura et al. 2003a; Lakatos et al. 2005) are conserved across species. Functional studies on the A755V variant co-occurring with the 1007fs mutation have shown a profound impairment of MDP responses (van Heel et al. 2006). In the present study, the allele frequencies of the A755V variant were higher in CD patients than in the background population, but this difference was not statistically significant. A higher 755V frequency in CD patients than controls has also been reported in another study (Lesage et al. 2002), but the number of mutated chromosomes was also relatively low. For the R791Q variant found in one CD patient no functional data are available at the moment.

The present data from structural analysis of the CARD15 gene support the idea that rare CARD15 mutations account only a small fraction of the susceptibility to CD. It is still
possible, however, that there are regulatory mutations in the 5’UTR or intronic areas of the gene and they contribute to the genetic susceptibility to CD.

The present study, using criteria based on the biochemical nature of the amino acid changes, allele frequencies in case-control analyses, and functional data available from the present and previous studies, suggests that 3% of Finnish CD patients carried two disease-susceptibility mutations while 16% carried one such mutation. These figures are much lower than those reported in the European collaboratory study where 17% of CD patients carried two mutations and 50% carried one mutation (Lesage et al. 2002). These results, combined with similar findings in Scotland (Arnott et al. 2004) and Scandinavia (Hampe et al. 2002b; Vind et al. 2005; Medici et al. 2006; Torkvist et al. 2006; Ernst et al. 2007), suggest the existence of other contributing susceptibility loci in CD.

1.1 CARD15 mutation frequency in the background population

The present study on CARD15 mutation frequency in healthy controls is the largest of its type to date. Data were obtained from 15 IBD research groups on three continents. It should be emphasized that these population samples were recruited by different methods, and most of them were not well characterised. The major bias that might affect the conclusions is the possibility of including actual CD patients in the background population. Furthermore, CD may not have developed yet in the youngest population controls screened.

Results from the present study and the literature show that in Caucasians, large differences are observed in the frequency of the three common mutations among apparently healthy individuals. In Europe, the CARD15 mutation frequency is relatively low in northern parts including Scotland and Scandinavia (Medici et al. 2006; Torkvist et al. 2006) and the highest in central European countries Ireland (Bairead et al. 2003), England, Belgium, France, Croatia (Cukovic-Cavka et al. 2006), Germany, and Hungary (Lakatos et al. 2005). In southern Europe, including Spain (Mendoza et al. 2003; Nunez et al. 2004), Portugal (Ferreira et al. 2005), and Italy, the mutation frequency is intermediate with the exception of Greece where it is compatible with central Europe (Gazouli et al. 2005b). These data are in accordance with more general observations of north-south gradients for genetic polymorphisms through Europe that are similar to the routes taken by Neolithic farmers from the Levant 10 000 years ago (Barbujani and Bertorelle 2001).

The CARD15 mutation frequency, however, is very low in Turkey (Uyar et al. 2004) and Israeli Arabs (Karban et al. 2005), whereas in Tunisia and Sephardic Jewish populations
from North Africa the \textit{CARD15} mutation frequency is somewhat higher (Tukel \textit{et al.} 2004; Zouiten-Mekki \textit{et al.} 2005). In North America, the mutation frequencies are similar to those observed in central Europe except for Quebec, which has a lower frequency. Finally, in the Pacific, the frequency is similar to southern Europe (Gearry \textit{et al.} 2006). Altogether, these data demonstrate that a large geographic heterogeneity exists in Caucasians and is most likely due to population history, genetic drift, admixture, and migrations.

The large differences observed in \textit{CARD15} mutation frequency between populations could indicate that it reflects the incidence of CD. Comparison of incidences and mutation prevalence could not, however, show a clear relationship. For example, in Finland CD incidence is relatively high (5.6/100 000) but the mutation frequency is low, while in Italy the situation is the opposite, having high mutation frequency with relatively low incidence (2.1/100 000). Lack of a clear relationship between mutation frequency in the background population and CD incidence indicate that \textit{CARD15} is only one of many different factors contributing to disease development and is not alone sufficient to influence disease incidence. Results from double mutation carriers indicate that the majority of mutation carriers are healthy and disease penetrance is limited, as expected for complex genetic diseases. This conclusion is also in accordance with incomplete disease concordance in monozygotic twins carrying the \textit{CARD15} mutations (Jess \textit{et al.} 2005).

1.2 Genotype-phenotype correlations of \textit{CARD15} mutations

Previous studies have consistently shown association between the \textit{CARD15} mutations and an ileal location of CD. In the present study, the same association was confirmed in the Finnish CD material. It was also seen that the association was dependent on the occurrence of the 1007\textit{fs} mutation. Strictureing and/or penetrating phenotypes have also been reported to associate with the common \textit{CARD15} mutations (Abreu \textit{et al.} 2002; Hampe \textit{et al.} 2002b; Lesage \textit{et al.} 2002; Brant \textit{et al.} 2003b; Economou \textit{et al.} 2004; Heresbach \textit{et al.} 2004; Oostenbrug \textit{et al.} 2006b; van der Linde \textit{et al.} 2007). This association was also observed in the present study: patients with two mutations more frequently had complicated phenotype than patients without or with one mutation.

In some studies, carrier status for two mutations has shown to associate with an earlier age at onset of CD (Lesage \textit{et al.} 2002; Brant \textit{et al.} 2003b; Newman \textit{et al.} 2004; Annese \textit{et al.} 2005a; Oostenbrug \textit{et al.} 2006b). In the present CD cohort, two sibling pairs and one sporadic patient, who were homozygous for the 1007\textit{fs} mutation, also had earlier ages at onset. In addition, these patients had more severe disease as indicated by occurrence of a
stricturing phenotype. This is in accordance with the study of Lesage and colleagues (Lesage et al. 2002), who showed that patients with two mutations had stricturing phenotype more often than patients without mutation.

2. GENETIC VARIATION IN OTHER IBD CANDIDATE GENES

2.1 Positive association with IL23R, TNFRSF1A, and HLA-DRB1*0103 allele variants

Since the original report concerning the IL23R gene with IBD and especially with CD susceptibility (Duerr et al. 2006), three other GWA studies have confirmed this finding (Consortium 2007; Libioulle et al. 2007; Raelson et al. 2007). Many replication studies have also reported association to both CD and UC (Buning et al. 2007; Cummings et al. 2007a; Glas et al. 2007; Oliver et al. 2007; Roberts et al. 2007; Tremelling et al. 2007; Weersma et al. 2008) while some have concentrated on CD only (Borgiani et al. 2007). In the present study, the involvement of the IL23R gene only with CD susceptibility was confirmed. In comparison to previous studies, a slightly weaker association to CD was obtained, possibly due to the facts that the present CD cohort was somewhat smaller and unselected instead of only ileal CD cases were examined.

The protective coding variant (R381Q) of the IL23R gene has shown the strongest association in most studies although the minor allele frequency is very low in all populations. The frequency of this protective variant was even lower in the present IBD cohort, and only six CD patients were heterozygous for the minor allele A. These patients shared no common clinical characteristics. Duerr and colleagues found an IL23R haplotype protective against CD (Duerr et al. 2006). It is also of note that the protective haplotype in the Finnish CD patients (CCTGATCG) differed by only one nucleotide from the haplotype in North American Caucasian CD patients (CGTGATCG). In addition, a significant association of two IL23R SNPs with stricturing behaviour of CD was demonstrated, and three SNPs were also associated with longer disease duration, both of which have only been reported in the present study.

In our previous genome-wide linkage scan, the highest two-point NPL score of 2.15 for IBD and the second highest NPL score of 2.00 for UC were observed on chromosome 12p13 (Paavola-Sakki et al. 2003). This chromosomal region also harbours the TNFRSF1A gene. When UC families were stratified according to their age at onset (≤24 years at diagnosis), the two-point NPL score reached 2.72. The present study shows that the TNFRSF1A gene variants A36G and IVS6+10A→G were significantly associated with the familial, but not sporadic, form of UC. Evidence was also obtained suggesting that the
same alleles were markers of ileocolonic involvement in CD. Other investigators have previously reported association between the TNFRSF1A A36G variant and CD as well as a negative association with stricturing phenotype (Waschke et al. 2005). Association of the same variant with the pancolitis phenotype of UC has been observed in the Belgian population (Pierik et al. 2004). No association between the A36G variant and IBD in Japanese population has been observed (Sashio et al. 2002).

The HLA-DRB1*0103 allele is very rare with less than 2% frequency in Caucasian populations but nevertheless it is the most replicated HLA class II allele in UC susceptibility (Ahmad et al. 2006). The association between the HLA-DRB1*0103 allele and UC was identified for the first time in 1996 (Satsangi et al. 1996b), and with CD a few years later (Trachtenberg et al. 2000). The present study confirms the association of this rare allele with UC among Finnish IBD patients, and the trend was similar for CD. The most consistent phenotypic findings among HLA-DRB1*0103 positive UC patients have been extensive colitis and severe disease as defined by the need for colectomy (Satsangi et al. 1996b; Roussomoustakaki et al. 1997; Bouma et al. 1999; Ahmad et al. 2003; Yamamoto-Furusho et al. 2003). In HLA-DRB1*0103 positive CD patients, a colonic location has been the most consistent finding (Silverberg et al. 2003; Fernandez et al. 2004; Newman et al. 2004). Extensive disease and need for colectomy as well as longer disease duration were more common in the present series of HLA positive UC patients but probably because of the overall rarity of this allele these phenotypic parameters were not statistically significant. In addition, all three CD patients carrying the HLA-DRB1*0103 allele had colonic involvement. These data strongly link the HLA-DRB1*0103 allele to inherited susceptibility of colonic inflammation, regardless of the exact subtype of IBD.

2.2 No association with TRAF6, ATG16L1, DLG5, ABCB1, TLR4, SLC22A4, SLC22A5, or the chromosome 5 haplotype

Only one intronic variant was detected upon TRAF6 gene screening. This variation, containing either seven or eight thymine (T) bases close to the intron 3/exon 4 boundary, was of potential interest because variable lengths of poly(T) tracts have been shown to associate with other conditions. For example, variable lengths of poly(T) tracts present close to the intron 8 splice branch/acceptor site result in alternative splicing of the human cystic fibrosis transmembrane conductance regulator (CFTR) mRNA (Chu et al. 1991; Chu et al. 1993; Chillon et al. 1995). In addition, mutations at 11 poly(T) repeat in intron 4 of the DNA repair protein MRE11 gene have been reported to cause aberrant splicing of mRNA and the generation of truncated proteins (Giannini et al. 2002). The opposite has been seen in the human interferon gamma (IFNG) gene where the different lengths of
poly(T) close to the splice acceptor site of intron 1 did not alter gene expression levels (Bream et al. 2000). Since the present study failed to show any evidence of disease association with the TRAF6 poly(T) tract, there was no point to carry on with mRNA splicing experiments.

This study was unable to confirm previous findings on the associations of ATG16L1, DLG5, ABCB1, or TLR4 variants, or the chromosome 5 risk haplotype including SLC22A4 and SLC22A5, with IBD. It is, however, possible that the number of patients studied was too small to detect an association, even if it is present in Finns. A phenotypic analysis showed some evidence for association between the TLR4 D299G polymorphism and extensive UC, but due to the lack of association in case-control analysis the significance of this finding remains obscure.
CONCLUDING REMARKS AND FUTURE PROSPECTS

IBD is a common disease of industrialised and developed countries, and the incidence has been growing since the 1950s. Management of IBD is often difficult, with recurrent symptoms and significant morbidity, and therefore treatment and hospitalisation are frequently associated with high cost and significant use of human resources.

One of the main goals was to assess the role of the \textit{CARD15} gene in a Finnish CD cohort. Three common \textit{CARD15} mutations R702W, G908R, and 1007fs associate independently with CD susceptibility, and the present data showed that 16\% of CD patients carried one of the three common mutations. Only the 1007fs mutation associated significantly with CD. As a result of \textit{CARD15} screening, five novel amino acid changes (R38M, W355X, P727L, W907R, and R1019X) were found. Novel variants are most likely disease-causing mutations on the basis of \textit{in vitro} functional studies, cross-species comparisons, and the nature of the amino acid substitutions. At the moment, \textit{CARD15} genotyping does not help much in the establishment of diagnosis or selection of treatment, and thus it is not routinely recommended to screen CD patients for \textit{CARD15} mutations.

Allele frequencies of the three common \textit{CARD15} mutations showed remarkable differences in apparently healthy individuals across three continents. The proportion of double mutant carriers showed that CD penetrance is limited, as expected for complex disease. No clear relationship was observed between \textit{CARD15} mutation frequency and disease incidence in different countries, indicating that there must be other factors involved in CD susceptibility.

Investigation of some of the recently described candidate genes also revealed an association of \textit{IL23R} and the HLA-DRB1*0103 allele with CD and UC susceptibility, respectively. In addition, an association between the \textit{TNFRSF1A} gene and familial UC was identified.

This is the first molecular genetic study concerning IBD in Finland. During the last decade there has been remarkable progress in the field of IBD genetics, and it is very likely that additional susceptibility genes will soon be found. High density mapping of potential loci identified from the Finnish genome scan and the conducting of a GWA study on Finnish UC patients might help to identify novel susceptibility genes. Identification of new genes may increase our understanding of the aetiology of IBD and could eventually improve diagnosis and assist in tailoring treatments for the disease.
ACKNOWLEDGEMENTS

The current study was carried out at the Research Program for Molecular Medicine, Biomedicum Helsinki and at the Department of Medicine, University of Helsinki during the years 2002-2008. The research project was part of the Centre of Excellence in Complex Disease Genetics (years 2006-2011) of the Academy of Finland. I wish to express my gratitude to professors Reijo Tilvis and Olavi Ylikorkala, the former and present heads of the Institute of Clinical Medicine, and professors Kimmo Kontula and Vuokko Kinnula, the former and present heads of the Department of Medicine, for providing excellent research facilities.

This study was financially supported by the Special State Share of the Helsinki University Central Hospital, the Research Funds of the University of Helsinki, and the Clinical Chemistry Research Foundation.

I am deeply indebted to my supervisors Dr Paulina Paavola-Sakki and professor Kimmo Kontula for guiding me through this research project. Paulina, you have been supportive when times have been rough, and I have enjoyed your company at our mutual IBD Consortium meetings. Kimmo, you are always so chronically interested in science and, in addition, you are an incredible scientific writer which never stops amazing me.

I am very grateful to docents Irma Järvelä and Kaija-Leena Kolho for carefully reviewing my thesis and for the most valuable comments and constructive criticism. Jodie Painter is acknowledged for thoroughly revising the language of my thesis.

I am indebted to our clinicians docents Martti Färkkilä, Leena Halme, and Tiina Heliö as well as Dr Ulla Turunen whose role in my thesis has been invaluable. This thesis would have not been possible without your effort collecting the patient material. Leena and Ulla are also acknowledged for reviewing the clinical part of my thesis. I am very grateful for the patients and families who have participated in the present study.

I wish to thank all my collaborators. Docent Heikki Repo and Eine Virolainen, at the Department of Bacteriology and Immunology, University of Helsinki, are acknowledged for invaluable collaboration. You also made me feel very welcome to your lab. My very warm thanks go to other collaborators docent Päivi Saavalainen and Elisabet Einarsdottir at the Research Program for Molecular Medicine as well.

Raija Selivuo is acknowledged for kind help with numerous practical matters.
I would like to thank Saara Nyqvist, Tuula Soppela, and Susanna Tverin for excellent technical assistance. In addition, I would like to acknowledge Tuula and Susanna for the friendship we share.

I have been honored to have the opportunity to work with many wonderful colleagues and friends. When I started at Kimmo’s lab, my closest colleagues were Heidi Fodstad, Päivi Forsblom, and Kaisa Valli-Jaakola. I have spent really good and memorable times with you at work and outside the work. Tuula Hannila-Handelberg, Annukka Marjamaa, Annukka Lahtinen, and Timo Suonsyrjä, I would like to wish all the best for your own thesis projects. This feeling is really something it is worth of doing. Present and former members of the group, Timo Hiltunen, Jukka Lehtonen, Marika Lilja, Maiju Merisalo, Laura Oksanen, Kristian Paavonen, Kirsi Paukku, and Kirsi Pippio are remembered with warmth. Kati Donner, the last but not least of our group member, you have always been there for me with both work and personal matters. I am truly happy we have made such a good friendship in a short period of time.

I wish to thank people from professor Aarno Palotie’s lab: Kirsi Alakurtti, Verneri Anttila, Eija Hämäläinen, Mari Kaunisto, Riitta Sallinen, Annika Sarahonka, Päivi Tikka-Kleemola, and docent Maija Wessman. Maija, you are the one who introduced me into this field and I am really grateful for that. I would like to mention Paula Kokko, Merja Lahtinen, Susanna Mehtälä, and Jaana Valkeapää for the pleasant company at coffee breaks.

I would like to thank my friends for endless support during this journey. The deepest gratitude goes to my parents Ulla and Pentti for continuous love and support. You have always trusted in me and let me do my own decisions in life. My sister Johanna has also been invaluable to me. Mikko’s parents Pirjo and Esko as well as his sister Paula along with her family have given me memorable moments outside the work, especially on our mutual trips abroad.

Finally, my dear life-companion Mikko, you have been there almost from the very beginning and stayed throughout this project. I am really grateful that you gave me all the time I needed at the end of this project. I am glad to have you in my life!

Helsinki, April 2008
REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


