Interplay of genetic and environmental triggers in intestinal inflammation: Genetics and transcriptomics in celiac disease and inflammatory bowel disease

AMARJIT SINGH PARMAR
Research Programs Unit, Immunobiology
University of Helsinki, Helsinki, Finland

ACADEMIC DISSERTATION
To be presented with the permission of the Faculty of Medicine, University of Helsinki, for public examination in Lecture Hall 1, Haartman Institute, Haartmaninkatu 3, Helsinki, on August 2nd 2013, at 12 noon

Helsinki 2013
TO MY FAMILY

Our greatest weakness lies in giving up. The most certain way to succeed is always to try just one more time.
Thomas A. Edison

If you're going through hell, keep going.
Winston Churchill
Table of Contents

List of original publications .................................................................................................................... 6
Abbreviations ........................................................................................................................................ 7
Abstract ................................................................................................................................................. 9
Introduction .......................................................................................................................................... 11
Review of the literature .......................................................................................................................... 12
1. The immune system at a glance ....................................................................................................... 12
   1.1. Innate immunity ............................................................................................................................. 12
   1.2. Adaptive immunity .......................................................................................................................... 13
      1.2.1. Th1 and Th2 cells and their role in autoimmune diseases .......................................................... 13
      1.2.2. Th17 cell subset and autoimmunity .......................................................................................... 14
      1.2.3. Regulatory T cells, Tr-1 cells and Th9 cells ........................................................................... 15
      1.2.4. Follicular helper T cells .......................................................................................................... 15
2. Celiac disease .................................................................................................................................... 16
   2.1. Epidemiology of celiac disease ....................................................................................................... 16
   2.2. Clinical features .............................................................................................................................. 17
   2.3. Genetics of celiac disease ............................................................................................................... 18
      2.3.1. Linkage studies in celiac disease ............................................................................................... 19
      2.3.2. Association studies in celiac disease ......................................................................................... 20
   2.4. Celiac disease pathogenesis- the role of innate and adaptive immunity ........................................ 25
   2.5. Diagnosis and treatment ................................................................................................................ 29
   2.6. Celiac disease and associated disorders ......................................................................................... 30
   2.7. Extraintestinal manifestation of celiac disease- dermatitis herpetiformis ..................................... 31
      2.7.1. Diagnosis and treatment of dermatitis herpetiformis ................................................................. 32
      2.7.2. Dermatitis herpetiformis and associated diseases ..................................................................... 33
3. Inflammatory bowel disease .............................................................................................................. 33
   3.1. Crohn’s disease (CD) ...................................................................................................................... 34
   3.2. Ulcerative colitis ............................................................................................................................ 35
   3.3. Clinical features of IBD ................................................................................................................... 37
   3.4. Epidemiology of IBD ...................................................................................................................... 39
   3.5. Genetics of IBD ............................................................................................................................... 41
      3.5.1. Family studies in IBD ................................................................................................................. 41
      3.5.2. Twin studies ................................................................................................................................ 42
      3.5.3. Linkage studies in IBD ................................................................................................................. 43
      3.5.4. Association studies in IBD ......................................................................................................... 45
   3.6. Pathogenesis of IBD ....................................................................................................................... 47
   3.7. Diagnosis and treatment ................................................................................................................ 48
4. Gut microbiota, immune system and susceptibility to different diseases ........................................49
5. Transcriptomics: scope and methods ................................................................................................54
6. Several risk genes are shared among different immunological diseases ........................................55

Aims of this study ..................................................................................................................................56

Material and methods ..........................................................................................................................57

1. Digestion of gliadin, bovine serum albumin and lactalbumin ..........................................................57
2. Caco-2 cell cultures and stimulations ..............................................................................................57
3. Cell collection and RNA extraction ................................................................................................58
4. Microarray hybridizations and data analysis ....................................................................................58
5. Quantitative real-time polymerase chain reaction (qRT-PCR) ..........................................................59
6. DNA extraction ..................................................................................................................................60
7. Marker selection ................................................................................................................................60
8. Study subjects ....................................................................................................................................60
9. Genotyping .......................................................................................................................................62
10. Power calculations ..........................................................................................................................64
11. Data analysis ...................................................................................................................................64

Results ...................................................................................................................................................66

1. Study I: The effect of gliadin peptides on Caco-2 cell gene transcription ........................................66
1.1. Differential gene expression in Caco-2 cells after gliadin treatment ............................................66
1.2. Validation of differentially expressed genes by qRT-PCR ............................................................66
2. Study II: Association of FUT2 with CelD, DH and IBD in the Finnish population ............................68
3. Study III: Association of CelD risk genes with IBD and with CD/UC subphenotypes ........................69
3.1. Association of CelD risk loci with CD and/or UC in the Finnish population ................................70
3.2. Association of CelD risk loci with subphenotypes of CD and/or UC in the Finnish population ......71
3.3. Meta-analysis of Finnish and Swedish UC datasets ......................................................................71

Discussion .............................................................................................................................................74

1. Modest effect of gliadin peptides and novel, non-enzymatic effects of pepsin and trypsin on Caco-2
gene transcription .................................................................................................................................74
2. Association of rs601338-A-FUT2 with CelD, DH and IBD in the Finnish population ........................81
3. Several CelD susceptibility genes also affect the risk for IBD ..........................................................83

Conclusions and future directions ........................................................................................................87

Acknowledgements ..................................................................................................................................89

References ............................................................................................................................................91
List of original publications

This thesis is based on the following original publications. The publications are referred to in the text by Roman numerals (I-III)


The original publications have been reproduced with permission from the copyright holders.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGA</td>
<td>anti-gliadin antibodies</td>
</tr>
<tr>
<td>APCs</td>
<td>antigen presenting cells</td>
</tr>
<tr>
<td>ARA</td>
<td>anti-reticulin antibodies</td>
</tr>
<tr>
<td>ATG2A</td>
<td>anti-tissue transglutaminase 2 antibodies</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CD</td>
<td>Crohn’s disease</td>
</tr>
<tr>
<td>CD4</td>
<td>cluster of differentiation 4</td>
</tr>
<tr>
<td>CD8</td>
<td>cluster of differentiation 8</td>
</tr>
<tr>
<td>CeD</td>
<td>celiac disease</td>
</tr>
<tr>
<td>DH</td>
<td>dermatitis herpetiformis</td>
</tr>
<tr>
<td>EAE</td>
<td>experimental autoimmune encephalomyelitis</td>
</tr>
<tr>
<td>EGFR</td>
<td>epidermal growth factor receptor</td>
</tr>
<tr>
<td>EMA</td>
<td>anti-endomysium autoantibodies</td>
</tr>
<tr>
<td>ESPGHAN</td>
<td>European Society of Pediatric Gastroenterology, Hepatology and Nutrition</td>
</tr>
<tr>
<td>FUT2</td>
<td>fucosyltransferase 2</td>
</tr>
<tr>
<td>GF</td>
<td>germ-free</td>
</tr>
<tr>
<td>GFD</td>
<td>gluten-free diet</td>
</tr>
<tr>
<td>GI</td>
<td>gastrointestinal</td>
</tr>
<tr>
<td>GPCRs</td>
<td>G protein-coupled receptors</td>
</tr>
<tr>
<td>GSE</td>
<td>gluten-sensitive enteropathy</td>
</tr>
<tr>
<td>GWAS</td>
<td>genome-wide association study</td>
</tr>
<tr>
<td>HLA</td>
<td>human leukocyte antigen</td>
</tr>
<tr>
<td>IBD</td>
<td>inflammatory bowel disease</td>
</tr>
<tr>
<td>IBDU</td>
<td>IBD unclassified</td>
</tr>
<tr>
<td>ICOS</td>
<td>inducible costimulator</td>
</tr>
<tr>
<td>IELs</td>
<td>intraepithelial lymphocytes</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>interferon gamma</td>
</tr>
<tr>
<td>IL-15</td>
<td>interleukin-15</td>
</tr>
<tr>
<td>IL-2</td>
<td>interleukin-2</td>
</tr>
<tr>
<td>iT\textsubscript{regs}</td>
<td>induced regulatory T cells</td>
</tr>
<tr>
<td>LD</td>
<td>linkage disequilibrium</td>
</tr>
<tr>
<td>LPMNCs</td>
<td>lamina propria mononuclear cells</td>
</tr>
</tbody>
</table>
LPP  lim domain-containing preferred translocation partner in lipoma
LPS  lipopolysaccharide
MAPK  mitogen-activated protein kinase
MDDCs  monocyte derived dendritic cells
MHC  major histocompatibility complex
NFkB  nuclear factor kappa-B
nTregs  naturally occurring regulatory T cells
OR  odds ratio
PAR2  proteinase-activated receptor-2
PT  pepsin and trypsin blank control
PT-BSA  pepsin and trypsin digested bovine serum albumin
PT-G  pepsin and trypsin digested gliadin
PT-L  pepsin and trypsin digested lactalbumin
RCD  refractory celiac disease
REL  v-rel avian reticuloendotheliosis viral oncogene homolog
SCFAs  short chain fatty acids
SH2B3  SH2B adapter protein 3
sIgA  secretory IgA
SLE  systemic lupus erythematosus
SNP  single nucleotide polymorphism
T1D  type-1 diabetes
TAGAP  T-cell activation GTPase activating protein
Tfh  follicular helper T cells
TFR  transferrin receptor
TGF-β  transforming growth factor beta
Th1  T helper cell type 1
Th17  T helper cell type 17
Th2  T helper cell type 2
TJ  tight junction
TNF  tumor necrosis factor
Tregs  regulatory T cells
tTG  tissue transglutaminase
UC  ulcerative colitis
Abstract

Celiac disease (CelD) and inflammatory bowel disease (IBD) [comprising Crohn’s disease (CD) and ulcerative colitis (UC)] are chronic inflammatory diseases of the gastrointestinal (GI) tract. In CelD, the disease triggering environmental factors are gluten peptides, derived from wheat, rye or barley. The genes within the human leukocyte antigen (HLA) region on chromosome six encoding HLA-DQ2 and –DQ8 heterodimers are strongly associated with CelD. However, in both IBD and CelD, non-HLA genes have been identified and some genes confer susceptibility to both CelD and IBD suggesting shared genetics in disease etiology. No definite environmental factors in IBD are known, but IBD is thought to be precipitated by inappropriate innate and adaptive immune response towards host gut microbiota. The fucosyltransferase 2 (FUT2) gene controls the expression of A-, B- and H- blood group antigens in mucus and other body secretions. Homozygosity for a nonsense mutation (rs601338G>A) in the FUT2 gene leads to the absence of ABH blood group antigens (non-secretor status). The secretor status affects the composition of gut microbiota which is important for proper development of the mucosal immune system. Thus, genetic variations in this gene may confer susceptibility to CelD and IBD. This study was conducted to identify genes which are affected by environmental factors (gliadin peptides) and to identify genes which affect environmental factors (gut microbiota) and also to explore the shared genetic background of CelD and IBD.

In Study I, the DNA microarray chips were used to study the effects of pepsin and trypsin (PT) digested gliadin (PT-G) and a synthetic gliadin peptide p31-43 on Caco-2 cell gene transcription. A blank PT control served as a negative control. Compared to the untreated medium control (MED-CTL), the expression of 1705 and 211 probes was affected by PT-G, and p31-43 respectively. The differential expression of three PT-G affected genes was confirmed by qRT-PCR. However, these genes were also affected by PT containing negative controls used in the study, suggesting a novel, non-enzymatic role for pepsin- and trypsin- derived peptides in Caco-2 gene regulation. The expression of the two genes differentially expressed after treatment with gliadin p31-43 peptide was also verified by qRT-PCR. No difference in gene expression in p31-43 peptide treated cells compared to the untreated control cells or the negative control was observed. This study concludes that the effects of gliadin peptides on enterocyte gene expression may be marginal. The presence of other cell types or chemical mediators may be required for the previously reported effects of gliadin in CelD.

In Study II, the association of rs601338-A-FUT2 with CelD, CelD patients with dermatitis herpetiformis (DH) denoted (CelD + DH), CD, UC and IBD was investigated in the Finnish population. The SNP rs601338 was genotyped in 1045 controls, 943 CelD patients, 119 DH patients, 303 CD
patients and 545 UC patients. The association analyses were performed using PLINK version v1.07. The \textit{FUT2} non-secretor genotype (rs601338-AA) was associated with increased risk for CelD. In addition, rs601338-A showed a nominal recessive association with CelD. The rs601338 SNP did not associate with DH, but showed significant association with CelD + DH. The rs601338 SNP did not associate with CD, but rs601338-A showed significant association with UC under the dominant association model. It is concluded that the \textit{FUT2} non-secretor status (rs601338-AA) is associated with increased risk for CelD but not IBD in the Finnish population.

In Study III, the association of novel CelD risk markers with IBD in the Finnish and Swedish populations was investigated. Forty-five CelD associated SNPs were genotyped in 699 Finnish IBD (CD: 240 and UC: 459) and 2482 control subjects. The Swedish dataset comprised 923 UC patients and 341 controls. The association of these SNPs with different subphenotypes of IBD was also investigated. Meta-analyses on Finnish and Swedish UC datasets were also performed. Several CelD risk SNPs were associated with CD and/or UC. Six SNPs each were associated with CD and UC and two SNPs (rs1893217-\textit{PTPN2} and rs4819388-\textit{ICOSLG}) were associated with both forms of IBD. Seven CelD risk alleles were protective (OR < 0.90) and 23 risk alleles conferred higher risk (OR > 1.10) for CD or UC. Twenty-two and 17 SNPs associated with a CD- or UC-specific subphenotype and 11 SNPs associated with both CD and UC subphenotypes. The SNP rs6974491-A-\textit{ELMO1} showed strong evidence for association with early-onset UC (onset age: 0-18 years) whereas rs2298428-T-\textit{UBE2L3} associated with early-onset CD (onset age: 0-18 years) and also with non-stricturing non-penetrating phenotype of CD. In the Finnish and Swedish UC meta-analysis, seven SNPs were significantly associated with UC. In the Finnish-Swedish UC subphenotype meta-analysis, nine SNP markers associated with one or more UC subphenotypes. The SNP rs1738074-A-\textit{TAGAP} showed an association with familial UC whereas rs4819388-G-\textit{COSLG} associated with sporadic UC. The CelD risk SNP rs6974491-A-\textit{ELMO1} associated with early-onset UC. Based on these results, it is thus concluded that several CelD risk SNPs are associated with CD and/or UC or their associated subphenotypes in the Finnish and Swedish populations.
Introduction

Celiac disease (CelD) and inflammatory bowel disease (IBD) are inflammatory diseases of the gastrointestinal (GI) tract which can affect an individual of any age. Both are regarded as complex diseases because of the involvement of multiple genes and environmental factors in disease onset and progression. The complexity of the disease increases several fold when unequal contribution of genes and/or environmental factors over a long time is taken into account. The environmental trigger in CelD is dietary gluten, whereas environmental factors which trigger the onset of IBD are largely unknown. The use of different genetic approaches (linkage studies and association studies) has led to the identification of several risk loci in CelD and IBD (Koskinen et al. 2008, Duerr et al. 1998, Hampe et al. 1999b, King et al. 2000, Dubois et al. 2010, Jostins et al. 2012). Linkage studies are effective in identifying rare causal mutations which are enriched in the family, whereas association studies present an effective way to identify common genetic variants that are enriched in disease affected, but unrelated, individuals. However, considering the small effect size that many genes are thought to have in complex diseases, such an approach require a great number of samples. This limitation is usually overcome by combining datasets derived from different populations and performing a meta-analysis (Dubois et al. 2010, Anderson et al. 2011, Franke et al. 2010). Recent research suggests that different immunological diseases share susceptibility genes (Festen et al. 2011, Zhernakova et al. 2011, Coenen et al. 2009). Many of the shared genes have important functions in the regulation of immune responses in disease pathogenesis. Studies investigating the common genetic background of different immunological diseases have furthered our understanding of shared disease pathogenesis (Zhernakova, van Diemen & Wijmenga 2009, Cotsapas et al. 2011). Nevertheless, despite advances made in identifying susceptibility genes in CelD and IBD, no cure for these diseases is yet available and the symptoms are treated through different approaches.

The purpose of this thesis was to investigate interactions between genes and environment in CelD and IBD and to explore the shared genetic background between these diseases. How the environmental trigger in CelD (i.e. gluten) affects epithelial cell gene transcription was investigated. Furthermore, the association of a genetic polymorphism in the fucosyltransferase 2 (FUT2) gene with CelD and IBD in the Finnish population was tested. Shared genetics between CelD and IBD was also investigated by studying the association of novel CelD associated risk loci in IBD cohorts from Finland and Sweden.
Review of the literature

1. The immune system at a glance

In everyday life, the human body is exposed to numerous potentially harmful environmental factors, living or nonliving, which can cause serious pathologies. Our immune system is however apt at providing adequate protection from such elements. In healthy individuals the two branches of the immune system, the innate immune system and the adaptive immune system, work in concert to protect the body from various threats.

1.1. Innate immunity

The immediate protection against pathogens is provided by innate immunity mediated by natural killer (NK) cells, mast cells, eosinophils, basophils, neutrophils, macrophages and dendritic cells. These are also known as the cells of the innate immune system. The first line of defense is however provided by epithelial cells which line the body cavities. The epithelial cells are interconnected through tight junctions and form a physical barrier against foreign antigens. Macrophages are tissue resident phagocytic cells which can recognize, ingest and destroy pathogens. When macrophages, through their cell surface receptors, interact with pathogens they are induced to synthesize and secrete cytokines and other chemical mediators. These molecules create a state of inflammation at the infection site. Inflammation recruits other effector cells to the infection site to clear the infection and also serves as a physical barrier to prevent further spread of the infection (Janeway et al. 2001). Neutrophils are also phagocytic cells which are present in large numbers in the blood but migrate to the site of infection during an innate immune response. Immature dendritic cells are present in different tissues throughout the body and bear receptors that recognize common features of many pathogens (Janeway et al. 2001, Underhill, Goodridge 2012). Dendritic cells and neutrophils help in the phagocytosis of invading pathogens. The innate immune response is non-specific and responds to different pathogens in much the same way. It does not generate immunological memory and a response to recurrent infections with the same pathogen is always the same. Although the innate immune system is highly effective in preventing the establishment of an infection, it falls short some times. This is primarily due to different mechanisms the pathogens have evolved to evade detection or destruction by the innate immune system. Only when the innate immune system fails to control the infection is the adaptive immune response initiated.
1.2. Adaptive immunity

An adaptive immune response comprises T-cell mediated immunity and humoral immunity. In contrast to the quick but non-specific nature of the innate immune system, adaptive immunity is highly specific. It also generates immunological memory, thus subsequent responses to infection with the same pathogen are rapid. However, T cells are unable to bind and respond to pathogens on their own and require presentation of small pathogen-derived peptides (antigens) by other cells for their activation. Tissue dendritic cells are known as professional antigen presenting cells (APCs) which specialize in taking up antigens and activating naive T cells (Janeway et al. 2001). Although less potent, macrophages and B cells are also able to present antigens to T cells.

When dendritic cells recognize and ingest microbial antigens during innate immune response, they are stimulated to express co-stimulatory molecules on their cell surface and to migrate to local lymphoid tissues. Here, the microbial antigens are presented to antigen-specific naive T cells complexed with either the major histocompatibility complex (MHC) class I or class II molecules. The antigens that bind to MHC class I molecules are generally virus-derived cytosolic peptides or self-peptides. These antigens are presented to naive T cells which express co-receptor protein CD8 (CD8+ T cells) on their cell surface. The peptides generated in acidic endocytic vesicles are complexed with MHC class II molecules and presented to CD4 co-receptor expressing T cells (CD4+ T cells). The MHC region is highly polymorphic and during development T cells are selected for recognizing foreign peptides bound to self-MHC molecules only. An APC must express co-stimulatory molecules and present antigens complexed with MHC molecules to fully activate T cells. Antigen presentation by APCs in the absence of co-stimulatory molecules results in T-cell tolerance. After activation, CD8+ T cells differentiate into cytotoxic T cells and specifically target infected cells for killing. The CD4+ T cells however differentiate into type-1 T helper (Th1) cells, Th2 cells, Th9 cells, Th17 cells, regulatory T (T<sub>reg</sub>) cells or follicular helper T (Tfh) cells (Dardalhon et al. 2008, Vinuesa et al. 2005). All the subsets of CD4+ T cells participate in different aspects of the host adaptive immune system.

1.2.1. Th1 and Th2 cells and their role in autoimmune diseases

The differentiation of CD4+ T cells into functionally distinct Th1 and Th2 cells, each with a specific cytokine profile, was described by Mossman et al. (Mosmann, Coffman 1989) in 1989. When Th1 cells are stimulated they produce large amounts of interferon gamma (IFN-γ) and also tumor necrosis factor (TNF)-α and interleukin-2 (IL-2) (Dardalhon et al. 2008, Roberts-Thomson et al. 2011). The CD4+ Th1 cells are primarily involved in protection against intracellular pathogens through macrophage activation and facilitate delayed-type hypersensitivity reactions and the production of selected antibodies. In contrast, Th2 cells have a prominent role in allergic responses and are
characterized by the production of cytokines such as IL-4, IL-5, IL-13 and IL-25 (Dardalhon et al. 2008). Th2 cells participate in the elimination of parasites and extracellular pathogens through the induction of B-cell isotype switching to IgE and IgG and antibody production. The differentiation of Th1 cells is promoted by APC-derived IL-12 but inhibited by IL-4 which is a Th2 cell-specific cytokine. Similarly, IFN-γ inhibits the differentiation and effector functions of Th2 cells (Roberts-Thomson et al. 2011, Liblau, Singer & McDevitt 1995).

The Th1 subset of CD4+ T cells was long believed to be largely responsible for the development of many autoimmune diseases (Liblau, Singer & McDevitt 1995). Studies on animal models for human autoimmune diseases showed the presence of Th1 cell-specific cytokines in the affected tissue at the peak of the disease and the presence of Th2 cell-specific cytokines during the recovery phase (Liblau, Singer & McDevitt 1995, Khoury, Hancock & Weiner 1992). The ability of adoptive transfer of Th1 cells to induce the disease in genetically susceptible mice (Baron et al. 1993) and disease prevention by anti-INF-γ antibodies (Campbell et al. 1991) or by the systemic administration of IL-4 (Rapoport et al. 1993) further strengthened this notion. Subsequently, it was shown that the reduction of IFN-γ signaling or a lack of other molecules critical in Th1 cell differentiation does not protect mice from experimental autoimmune encephalomyelitis (EAE), an animal model for multiple sclerosis (MS), but results in a more severe disease (Dardalhon et al. 2008, Ferber et al. 1996). These studies suggested the existence of another type of effector T cell in organ-specific autoimmunity.

1.2.2. Th17 cell subset and autoimmunity

Interleukin-12 is composed of two subunits, p35 and p40, with the p40 subunit being shared with another cytokine, IL-23. The IL-23 cytokine consists of the IL-23p40 and IL-23p19 subunits and is produced predominantly by macrophages and dendritic cells. Cua et al. showed that mice which lack the IL-23 p40 or p19 subunits are resistant to EAE development, but those that lacked the IL-12 p35 subunit are not (Cua et al. 2003). They suggested that IL-23, rather than IL-12 or Th1 cells in general, is critical for EAE development and probably also for other autoimmune diseases. It was later found that in the presence of transforming growth factor beta (TGF-β) and IL-6, naive CD4+ T cells differentiate into a distinct cell type. This newly identified cell type, now known as Th17, expresses the IL-23 receptor and produces large quantities of IL-17, IL-17F, IL-6, IL-21, IL-22 and tumor necrosis factor (TNF) (Bettelli et al. 2006). Confirming the pathogenic nature of IL-23-dependent Th17 cells, it was shown that passive transfer of Th17 cells into naive mice causes EAE in these mice (Langrish et al. 2005). This effect could be partially blocked using anti-IL-17 antibodies and the authors argued that other factors induced by IL-23 (IL-6, IL-17F and TNF) are also important in EAE pathogenesis. Since their discovery, Th17 cells have been implicated in many autoimmune diseases including

1.2.3. Regulatory T cells, Tr-1 cells and Th9 cells

Another subset of Th1 cells, termed regulatory T (Tregs) cells, has been shown to have a reciprocal developmental relationship with Th17 cells (Bettelli et al. 2006). The Treg cells derive from CD4+ T cells in the presence of TGFβ alone, unlike Th17 cells which also require IL-6 (Jager, Kuchroo 2010). Regulatory T cells express the transcription factor FOXP3 and can be further subdivided into two groups, naturally occurring Tregs (nTregs) and induced Tregs (iTregs). Naturally occurring Tregs are generated in the thymus and iTregs differentiate from naive T helper cells in the periphery in the presence of TGFβ (Jager, Kuchroo 2010). Tregs are important in maintaining self-tolerance and in controlling the expansion and activation of autoreactive CD4+ T effector cells. Thus, Tregs play an important role in autoimmune pathogenesis. Furthermore, CD4+ T cells exposed to TGFβ and IL-27 differentiate into Tr-1 cells (Awasthi et al. 2007) while those exposed to TGFβ and IL-4 differentiate into Th9 cells (Soroosh, Doherty 2009). In mice, Tr-1 cells do not express Foxp3, but possess strong immunosuppressive properties and produce the cytokines IL-10 and IFN-γ (Jager, Kuchroo 2010). On the other hand, Th9 cells produce IL-9 and IL-10 cytokines and may play a role in defense against certain helminth infections (Jager, Kuchroo 2010) as well as in asthma and allergy (Soroosh, Doherty 2009).

1.2.4. Follicular helper T cells

An adaptive immune response comprises a T-cell mediated immune response and a B-cell mediated immune response. The CD4+ T cells and the CD8+ T cells are the dominating cell types in T-cell mediated immunity. In contrast, the humoral immune response which protects the extracellular spaces of the body is mediated by B cells. In the humoral immune response, B cell-produced antibodies cause the destruction of pathogens and prevent their spread to other parts of the body (Janeway et al. 2001). However, as T cells require help from APCs for activation, B cells require help from Th cells for activation and differentiation into antibody-secreting plasma cells. IFNγ produced by Th1 cells stimulates B cells to produce IgG2a class antibodies and Th2 cells-derived IL-4 promotes the production of IgE class antibodies. A distinct subset of CD4+ T cells, called follicular helper T (Tfh) cells, has been identified (Schaeerli et al. 2000). These cells are found in germinal centers, have a stable expression of chemokine receptor-5 (CXC5) and provide help to germinal center B cells (Vinuesa et al. 2005, Schaeerli et al. 2000). Experiments in mice demonstrate that Tfh cells have a distinct gene expression profile compared to Th1, Th2 and Th17 cells (Nurieva et al. 2008). The generation of Tfh cells is dependent on signal transducer and activator of transcription (STAT)-3
signaling and requires the presence of IL-6 and IL-21. Tfh cells have been implicated in autoimmune diseases, but their role is not entirely clear (Vinuesa et al. 2005, Weinstein, Hernandez & Craft 2012).

2. Celiac disease

Celiac disease (CelD) or gluten-sensitive enteropathy (GSE) is a chronic immune-mediated disease of the small intestine characterized by villous atrophy and crypt cell hyperplasia. CelD is caused by the ingestion of dietary gluten and affects genetically susceptible individuals. Patients suffering from CelD display a wide spectrum of symptoms, both intestinal and extraintestinal. The symptoms in CelD can be treated only by maintaining a strict life-long gluten-free diet (GFD).

2.1. Epidemiology of celiac disease

Although the prevalence of CelD is around 1% worldwide, it displays a great variation among European countries, ranging from 0.3% in Germany (Reilly, Green 2012) to 2.4% in Finland (Mustalahti et al. 2010). In Finland, a significant increase (from 1.03% to 1.99%) in CelD prevalence over a 20 year period has been reported (Lohi et al. 2007). In an Italian cohort (3483 study subjects, aged 12-65 years), the prevalence of CelD was found to be 5.7 per 1000 individuals (0.57%) (Volta et al. 2001). However, when the prevalence was studied after conditioning for age, prevalence in the age group 12-25 years (784 individuals) was highest at 10.2 per 1000 individuals (1.02%). In a study by Riestra et al. which included 1170 randomly selected subjects from an area in Northern Spain, the prevalence of CelD was reported to be 2.6 per 1000 (0.26%) (Riestra et al. 2000). In accordance with the high proportion of individuals with European ancestry in the United States, Australia and New Zealand, prevalence rates similar to that in Europe have been reported. The prevalence of CelD in the United States is close to 1% in the general population (Fasano et al. 2003) with more pronounced risk for first-degree and second-degree relatives of CelD patients. In large population based studies from Australia and New Zealand, the prevalence of CelD was reported to be 0.4% (Hovell et al. 2001) and 1.2% (Cook et al. 2000) respectively. Although CelD has previously been considered to be more widespread in Caucasians, it is now increasingly being recognized in Asia, the Middle East and North Africa. In the North Indian state of Punjab, the prevalence of CelD in 4347 school children aged 3-17 years was 0.3% (Sood et al. 2006). In a study performed in the UK, a higher incidence rate for UK-resident North Indians (6.9/100 000/year) was reported compared to Caucasians (2.5/100 000/year) living in the same area (Sher et al. 1993). Similar results have been reported in a study performed in Iran, which consisted of 2000 healthy blood donors representing the general population (Shahbazkhani et al. 2003). One person in every 166 individuals was predicted to have CelD. In a study carried out in two Egyptian pediatric cohorts, the prevalence of
CelD in one cohort (1500 children, seven months-18 years) was 0.53% (Abu-Zekry et al. 2008). The prevalence of CelD in the other group, (150 children, six months-13 years) admitted for diarrhea or failure to thrive, was 4.7%. In 250 children and adolescents with type-1 diabetes (T1D), the CelD prevalence was reported to be 6.4%. In North Africa, a CelD prevalence of 5.6% has been reported in Saharawi children (Catassi et al. 1999). There are reports suggesting that females are diagnosed two to three times more frequently compared to their male counterparts (Bai et al. 2005, Green et al. 2001) and that males have more severe symptoms at presentation. These studies suggest that CelD is no longer a disease of the West, but one affecting an equivalent number of people in other parts of the World as well.

CelD patients suffer from a wide spectrum of intestinal and extraintestinal symptoms which severely affects their quality of life. Both genetic predisposition and environmental factors are necessary for disease onset. Although gluten is the ultimate trigger for CelD onset, other environmental factors such as the duration of breast-feeding in childhood, timing of gluten introduction in an infant’s diet and amount of gluten when introduced as well as the composition of gut microbiota have been proposed to affect CelD risk (Ludvigsson, Fasano 2012). An increased risk for CelD has also been associated with childhood infections with Adenovirus 12 and Hepatitis C virus (Plot, Amital 2009). However, Myleus et al. reported that repeated infections, of any type, in the first six months of life significantly increased the risk for CelD in later life (Myleus et al. 2012). In strong contrast to this, it has also been suggested that early infections may have a protective role in CelD (Plot et al. 2009), which is in agreement with the hygiene hypothesis in autoimmune diseases (Rook 2012). According to the hygiene hypothesis, increased cleanliness in developed countries has resulted in diminished interaction between microbes and humans. This has had a negative impact on the development of an immune system tolerant to innocuous environmental antigens. To date, a life-long GFD remains the only treatment for CelD.

2.2. Clinical features

CelD was long considered a disease of the childhood characterized by malabsorption and steatorrhea (Andersen 1947). It is now clear that it can affect genetically susceptible individuals at any age. CelD patients suffer from a wide range of intestinal and/or extraintestinal symptoms. An individual may present with symptoms indicative of classical or non-classical CelD. Classical CelD is considered to arise in early infancy and is characterized by villous atrophy and crypt cell hyperplasia (Figure 1). The typical symptoms of classical CelD include diarrhea, steatorrhea, poor appetite, failure to thrive, muscle wasting, malnutrition and abdominal distension and pain (Young, Pringle 1971, D’Amico et al. 2005). Non-classical CelD is more prevalent among older children and adults.
These patients have a few or no intestinal symptoms reminiscent of classical Celiac Disease. Symptoms associated with non-classical Celiac Disease often include recurrent abdominal pain, nausea, vomiting, bloating, dental enamel defects and recurrent aphthous stomatitis (Bucci et al. 2006). These patients may also present with extraintestinal symptoms such as arthritis, neurological symptoms and anemia or the disease may be silent without any apparent symptoms (Ludvigsson et al. 2004, Branski, Troncone 1998). Such diverse symptoms sometimes lead to a delay in diagnosis. Undiagnosed and newly diagnosed Celiac Disease patients may suffer from anemia, which may be a result of poor iron absorption, folate deficiency or vitamin B12 deficiency (Tikkakoski, Savilahti & Kolho 2007, Halfdanarson, Litzow & Murray 2007). These patients recover from anemia by following a gluten-free diet in conjunction with iron supplements. Less than 5% of Celiac Disease patients develop refractory celiac disease (RCD). Such patients experience persistent symptoms, severe villous atrophy, do not respond to a gluten-free diet and may develop other complications later (Verbeek et al. 2008).

2.3. Genetics of celiac disease

The genetic architecture of an individual is just as important for Celiac Disease development as the disease triggering environmental factor, gluten. Genetic variants within the HLA locus on chromosome six are associated with increased risk for Celiac Disease (Sollied et al. 1989). The primary risk for developing Celiac Disease is associated with the HLA-DQ2 heterodimer which is composed of the protein products of the HLA-DQA1*0501 and HLA-DQB1*0201 alleles (Sollied et al. 1989). This heterodimer is present in about...
95% of CelD patients of European descent. The alpha- and beta chains of the DQ2 heterodimer can be encoded in cis (i.e. both on the same DRB1*03 (DR3) haplotype, also called the DQ2.5 haplotype) or in trans (i.e. with the DQA1*05 allele on DRB1*11, DRB1*12 or DRB1*13 haplotype and the DQB1*02 allele on DRB1*07 (DR7) haplotype) configuration (Spurkland et al. 1992, Mazzilli et al. 1992). The remaining 5% of CelD patients carry the HLA-DQ8 heterodimer (Spurkland et al. 1992). The DQ8 heterodimer is composed of proteins encoded by the HLA-DQA1*03 and DQB1*0302 alleles. The HLA-DQ2 and -DQ8 encoding genotypes are quite common (20%-30%) in the general population in Western countries (Sollid et al. 1989, Polvi et al. 1996, Cummins, Roberts-Thomson 2009). They have also been reported to occur at a relatively high frequency in Northern and Western Africa, the Middle East and Central Asia (Cummins, Roberts-Thomson 2009). In contrast, the HLA-DQ8 encoding genotypes have a worldwide distribution, whereas DQ2.5 is quite common in South and Central America and DQ8 is present in about 90% of Amerindian populations (Gujral, Freeman & Thomson 2012).

The concept of HLA-DQ2 gene dose effect in CelD is well established (Karinen et al. 2006, Congia et al. 1994). Vader and colleagues demonstrated that HLA-DQ2.5 homozygous individuals have the highest risk for CelD development followed by HLA-DQ2.2/2.5 heterozygous individuals (Vader et al. 2003). The HLA-DQ2 or -DQ8 heterodimers are required but not sufficient for CelD development suggesting the existence of other non-HLA disease predisposing genes. Further insight into the contribution of HLA- and non-HLA genes in CelD is provided by studies comprising monozygotic and dizygotic twin-pairs. The concordance rate between monozygotic twins is significantly higher (up to 75%) than in dizygotic twins (up to 11%) (Greco et al. 2002, Nistico et al. 2006). The concordance rate between monozygotic twins and HLA identical sibs is estimated to be 70% and 30% respectively and the risk contribution of the MHC region is suggested to be no more than 40% (Houlston, Ford 1996, Bevan et al. 1999). Genetic studies have been performed to identify non-HLA genes which confer risk for CelD development.

2.3.1. Linkage studies in celiac disease

Linkage studies are performed in families with at least two affected individuals, usually siblings, and the segregation of specific genetic markers (microsatellites or single-nucleotide polymorphisms (SNPs)) is followed within the family. SNPs are distributed over the whole genome and inherited by offspring from parents. The idea behind following certain SNPs in genome-wide linkage scans to identify disease associated risk genes is that SNPs and haplotypes close to a disease causing gene are likely to be inherited identically by the affected siblings or other relatives. As compared to free inheritance, the loci with an excess of sharing between relatives in all the tested families show
genetic linkage to the disease. Linkage studies have been used to study small genomic region harboring potentially interesting genes or at a genome-wide level. In candidate gene studies, there is some evidence or indication for the involvement of the gene in disease susceptibility. This approach has led to the identification of several risk loci which have been independently replicated in different populations. Multiple studies have reported linkage to 2q33 (CELIAC3) (Holopainen et al. 1999, Naluai et al. 2000, Popat et al. 2002b, Holopainen et al. 2004). Some studies have reported linkage to 5q31-33 (CELIAC2) (Koskinen et al. 2009b) and 19p13 (CELIAC4) (Koskinen et al. 2008, Van Belzen et al. 2003). However, in a study by Seegers et al., no evidence of linkage to CELIAC2 region was found in a Dutch cohort (Seegers et al. 2003). The 2q33 region harbors genes like CTLA4, ICOS and CD28, which have roles in T-cell mediated immunity and the MYO9B gene, located in the 19p13 chromosomal region, may be involved in the regulation of actin cytoskeleton or tight junctions. In contrast to the candidate gene studies, genome-wide studies are assumption free. A few chromosomal regions have been repeatedly highlighted in different genome-wide linkage scans, e.g. 5q31-33 (Greco et al. 1998, Naluai et al. 2001), 2q33 (King et al. 2000, Naluai et al. 2001), and 19p13 (Van Belzen et al. 2003, Popat et al. 2002a, van Belzen et al. 2004). These studies have been summarized in Table 1. The linkage approach has been quite successful in monogenic Mendelian diseases. However, results in polygenic complex diseases have been less fruitful as linkage scans in different populations have been somewhat contradictory.

Table 1. Linkage studies performed in celiac disease

<table>
<thead>
<tr>
<th>Locus/Chromosome</th>
<th>Candidate gene linkage studies</th>
<th>Genome-wide linkage studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>CELIAC2/5q31-33</td>
<td>(Koskinen et al. 2009b, Seegers et al. 2003)</td>
<td>(Greco et al. 1998, Naluai et al. 2001)</td>
</tr>
</tbody>
</table>

2.3.2. Association studies in celiac disease

In contrast to family-based linkage studies, association studies are performed on a group of unrelated disease affected individuals (referred to as cases) and healthy individuals (referred to as controls). This method compares allele frequency (or genotype distribution), in one or more SNPs (or
other genetic polymorphisms) at a given locus in the genome, between cases and controls. A SNP is said to be associated with a disease if a significant difference in allele frequency (or genotype distribution) is observed between cases and controls. Alleles that exist close to each other at a given genetic locus tend to be inherited together more often than is expected by chance. Such alleles are said to be in linkage disequilibrium (LD). SNPs (or other polymorphisms) that are in LD share common ancestry because they have not been separated by mutations or recombination events across generations (Gabriel et al. 2002). The strength of LD between two markers is assessed by $D'$ or $r^2$. A $D'$ or $r^2$ value equal to zero means that the markers are independent and a value equal to one means that the two markers coexist. If the rarer allele always occurs with one of the two alleles at the other marker, $D'$ is equal to one. In contrast, $r^2$ is equal to one if the allele frequencies between the two SNPs are identical. Thus, if two or more SNPs are in LD, it is enough to genotype one of these SNPs as it provides information on genetic variants at other loci in the LD group. In association studies, this method greatly reduces the number of SNPs required to be genotyped to cover common genetic variants at a given locus.

The **CELIAC3** locus identified in linkage scans was independently validated by association studies (Holopainen et al. 1999, Djilali-Saiah et al. 1998, Hunt et al. 2005). The association of SNPs in the **CELIAC2** and **CELIAC4** loci has been reported in some studies (Koskinen et al. 2009b, Van Belzen et al. 2003, Monsuur et al. 2005) but not in others (Koskinen et al. 2008, Giordano et al. 2006, Hunt et al. 2006, Ryan et al. 2005). The association approach to candidate gene studies has identified several other genomic regions harboring genes, among others **intracellular adhesion molecule-1** (ICAM-1) (Abel et al. 2006), **immunoglobulin G Fc receptor 2a** (FcgR2a) (Alizadeh et al. 2007), **IL-10** (Barisani et al. 2006) and **IL23R** (Nunez et al. 2008). Nevertheless, other studies have reported a lack of association between polymorphisms within the ICAM-1 (Dema et al. 2008) and IL23R genes (Weersma et al. 2008, Einarsdottir et al. 2009). A summary of the studies described here is given in Table 2.

In the first CelD genome-wide association study (GWAS) (Table 2), the association of 310,605 SNPs was tested in 778 CelD patients and 1,422 healthy individuals (van Heel et al. 2007). In addition to the HLA locus, this study identified a genomic region on chromosome four encompassing the **KIAA1109-TENR-IL2-IL21** genes. This finding was further validated in Dutch and Irish cohorts. Of the four genes in this region, the genes **IL2** and **IL21** are particularly interesting. The IL2 cytokine can stimulate T-cell activation and proliferation in an autocrine manner and IL21 enhances B cell, T cell and NK cell proliferation and IFN-$\gamma$ production. The GWAS by van Heel et al. was a landmark study which set the stage for other GWAS performed in different populations. One thousand twenty of
the most strongly associated SNPs in the GWAS by van Heel et al. (van Heel et al. 2007) were studied in a follow up association study (Hunt et al. 2008). This study consisted of independent CelD cases and healthy control subjects (cases: 1643, controls: 3406) from the UK, Ireland and the Netherlands. Twenty-one non-HLA SNPs from eight different chromosomal locations (1q31, 2q11-2q12, 3p21, 3q25-3q26, 3q28, 6q25 and 12q24) meeting a genome-wide statistical significance (P overall $< 5 \times 10^{-7}$) were reported in this study. Six of the loci identified harbor genes (1q31: regulator of G-protein signaling 1, RGS1; 2q11-2q12: interleukin-18 receptor accessory protein, IL18RAP; 3p21: chemokine receptor 3, CCR3; 3q25-3q26: interleukin-12A, IL12A; 6q25: T-cell activation GTPase activating protein, TAGAP and 12q24: SH2B adapter protein 3, SH2B3) which are involved in the regulation of immune responses. A marker (rs1464510) at 3q28 locus on chromosome three, containing the lim domain-containing preferred translocation partner in lipoma (LPP) gene was also associated with CelD. Although LPP is expressed at high levels in the small intestine, its function in CelD pathogenesis is not entirely clear. Findings from the first CelD GWAS and the follow-up association study by Hunt et al. (van Heel et al. 2007, Hunt et al. 2008) were followed up in an Italian cohort comprising 538 CelD patients and 593 healthy controls (Romanos et al. 2009). Six of the eight new CelD loci were replicated in this study and, interestingly, two loci (3p21-CCR3; 2q11-2q12-IL18RAP) showed no association with CelD in the Italian population. In contrast, another study confirmed the association of IL18RAP locus with CelD in a Hungarian cohort (cases: 607, controls: 448), but not in Finnish (cases: 844, controls: 698) or Italian (cases: 187, controls: 239) cohorts (Koskinen et al. 2009a). Results from these studies point towards the presence of population-specific differences in susceptibility genes. Trynka et al. studied the association of 458 SNPs that were only modestly associated with CelD in the GWAS by van Heel et al. (van Heel et al. 2007, Trynka et al. 2009) This study included samples from four different Caucasian populations (total cases: 2987, total controls: 5273) and identified two new (6q23.2-OLIG3, TNFAIP3 and 2p16.1-REL) risk loci. The protein products of TNFAIP3 and REL are involved in the regulation of the nuclear factor kappa-B (NFkB) pathway which plays a critical role in promoting immune and inflammatory responses. CelD is a complex disease and several genes with small effect size may also play a role in its predisposition. Discovering such genes requires large sample sizes.

A second GWAS was carried out in five European CelD case and control sample collections (total cases: 9,451 and total controls: 16,434) (Dubois et al. 2010). Such a study may not be able to identify population-specific risk loci due to genetic heterogeneity of the samples, but is likely to discover common risk genes with relatively low effect size. This GWAS identified several non-HLA genes with known functions in immune system regulation (Dubois et al. 2010). In addition to confirming association [with genome-wide significant evidence (P $< 5 \times 10^{-8}$)] with 13 previously identified non-
HLA CeLD risk loci (RGS1, REL, IL18RAP, ITGA, CTLA-4/ICOS/CD28, CCR1, CCR 3, CCR5, CCR9, IL-12A, LPP, IL-2/IL-21, TNFAIP3, TAGAP, SH2B3 and PTPN2), association with thirteen new loci (TNFRSF14, RUNX3, PLEK, CCR4, CD80, BACH2, THEMIS, ZMIZ1, ETS1, CIITA, CLEC16A and ICOSLG) was reported. Furthermore, another 13 loci (PARK7/TNFRSF9, NFIA, CD247, FASLG, TNFSF18, FRMD4B, IRF4, ELMO1, ZFP36L1, UBEEL3, TLR7/TLR8) were suggested to be associated with CeLD. Although this study brought the number of confirmed CeLD non-HLA risk loci to 26, these loci only explain a fraction of the disease heritability and most of the common and rare risk variants remained unknown. The immunochip consortium (Cortes, Brown 2011) was developed to identify additional rare and common risk variants and causal genetic polymorphisms in immunological diseases. In order to densely genotype loci associated with an immunological disease using the 1000 genomes project data (1000 Genomes Project Consortium et al. 2012) and other available disease-specific resequencing data, a custom Illumina Infinium High-Density array (called the immunochip) was developed. This consortium took into account that many risk loci are shared across different autoimmune or chronic immune-mediated diseases (Zhernakova et al. 2011, Coenen et al. 2009, Zhernakova, van Diemen & Wijmenga 2009, Zhernakova et al. 2007, Smyth et al. 2008). Thus, for the immunochip, the consortium selected 186 distinct loci containing genome-wide significant markers (P < 5 x 10^{-8}) from 12 immune-mediated diseases (UC, CD, T1D, SLE, rheumatoid arthritis (RA), psoriasis, primary biliary cirrhosis, MS, IgA deficiency, CeLD, ankylosing spondylitis and autoimmune thyroid disease). Using the immunochip, Trynka et al., densely genotyped 183 non-HLA risk loci in 12 041 CeLD patients and 12 228 healthy subjects and the association of 139 553 SNPs was studied (Trynka et al. 2011). SNPs at these loci have been reported previously to be associated with immune-mediated diseases. This study identified 13 novel risk loci (FASLG, STAT4, IRF4, PTPRK, EMLO1, PFKFB3/PRKCQ, POU2AF1/C11orf93, Treh/DDX6, ZFP36L1, CLK3/CSK, CIITA, UBASH3A, UBE2L3/YDJC and HFC1/TMEM187/IRAK) taking the total number of CeLD susceptibility loci to 40. The presence of certain variants within these regions inappropriately affect the innate and adaptive immune responses to gluten peptides ensued by a persistent intestinal inflammation. Inarguably, the immune system has a central role in disease progression and identifying factors that modulate the immune system is a key step in solving the CeLD puzzle.
Table 2. Association studies performed in celiac disease.

<table>
<thead>
<tr>
<th>Loci/Genes identified or investigated</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CELIAC2</strong> (5q31-q33)</td>
<td>(Koskinen et al. 2009b, Ryan et al. 2005)</td>
</tr>
<tr>
<td><strong>CELIAC3</strong> (2q33)</td>
<td>(Holopainen et al. 1999, Djilali-Saiah et al. 1998, Hunt et al. 2005)</td>
</tr>
<tr>
<td>FcgR2a (1q23.3)</td>
<td>(Alizadeh et al. 2007)</td>
</tr>
<tr>
<td>IL-10 (1q32.1)</td>
<td>(Nunez et al. 2008, Weersma et al. 2008, Einarsdottir et al. 2009)</td>
</tr>
<tr>
<td><strong>IL23R</strong> (1p31.3)</td>
<td>(van Heel et al. 2007)</td>
</tr>
<tr>
<td>RGS1 (1q31), IL18RAP (2q11-q12), CCR3 (3p21), IL12A (3q25-q26), TAGAP (6q25), SH2B3 (12q24)</td>
<td></td>
</tr>
<tr>
<td><strong>IL2-IL21</strong> (4q27), RGS1 (1q31), IL12A/SCHIP1 (3q25-q26), LPP (3q28), SH2B3/ATNX2 (12q24), TAGAP (6q25)</td>
<td>(Romanos et al. 2009)</td>
</tr>
<tr>
<td><strong>OLIG3-TNFAIP3</strong> (6q23.3), REL (2p16.1)</td>
<td>(Trynka et al. 2009)</td>
</tr>
<tr>
<td>TNFRSF14 (1p36.32), RUNX3 (1p36.11), PLEK (2p14-p13), CCR4 (3p22.3), CD80 (3q13.33), BACH2-MAP3K7 (6q15), PTPRK-THEMIS (6q22.33), ZMIZ1 (10q22.3), ETS1 (11q24.3), CIITA-SOCS1-CLEC16A (16p13.13), ICOSLG (21q22.3)</td>
<td>(Dubois et al. 2010)</td>
</tr>
<tr>
<td>FASLG (1q24.3), STAT4 (2q32.2-q32.3), IRF4 (6p25.3), PTPRK (6q22.33), EMLO1 (7p14.2-p14.1), PFKFB3/PRKCO (10p15.1), POU2AF1 (11q23.1), TREG/DDX6 (11q23.3), ZFP36L1 (14q24.1), CLK3/CSK (15q24.1), CIITA (16p13.13), UBASH3A (21q22.3), UBE2L3/YDJC (22q11.2-q13.1) and HCFC1/TMEM187/IRAK (Xq28)</td>
<td>(Trynka et al. 2011)</td>
</tr>
</tbody>
</table>

*underlined: no evidence for association*
2.4. **Celiac disease pathogenesis - the role of innate and adaptive immunity**

Already in the 1940s, the Dutch pediatrician Willem-Karel Dicke realized that there was something in wheat that was responsible for causing CelD (van Berge-Henegouwen, Mulder 1993) but only decades later gliadin peptides were implicated as the main disease trigger (Figure 2). Gluten is a storage protein present in food grains such as wheat, rye and barley. It comprises alcohol-soluble gliadins and alcohol-insoluble glutenins. In the stomach and small intestine, the proteolytic enzymes pepsin and trypsin, respectively, act upon these proteins. The high content of glutamine and proline in gliadin makes it highly resistant to degradation by proteolytic enzymes (Shan et al. 2002, Shan et al. 2005). Although these peptides evoke both the innate and adaptive immune systems, it is the onset of mucosal adaptive immunity to certain gliadin peptides (immunodominant peptides e.g. p57-68) which is detrimental in CelD. There are also cytotoxic gliadin peptides (p31-43 or p31-49) which trigger the innate immune system (Maiuri et al. 2003a) and induce enterocyte apoptosis (Maiuri et al. 2003a, Giovannini et al. 2003) and proliferation (Barone et al. 2011).

Moss *et al.* reported an increased rate of apoptosis in biopsy samples from untreated CelD patients compared to that seen in controls (Moss et al. 1996). Similarly, a significant increase in the number of proliferating cells in the crypts was observed in biopsies from active CelD patients. Both, apoptosis and proliferation rate decreased in these patients upon gluten withdrawal from diet demonstrating their gluten dependence. The involvement of the Fas-FasL pathway has been investigated using CelD biopsy samples. Maiuri *et al.* showed that *in vitro* gliadin challenge of biopsy samples from untreated CelD patients induces apoptosis via a Fas-dependent pathway (Maiuri et al. 2001) and IELs of more than 50% of untreated CelD patients express FasL. A coculture experiment using enterocytes and lamina propria mononuclear cells (LPMNCs) isolated from untreated CelD patients was performed (Ciccocioppo et al. 2001). It was shown that FasL expressing LPMNCs selectively target Fas expressing enterocytes and induce apoptosis. In the lamina propria, gliadin peptides (in particular the p31-43 peptide) induce the expression and secretion of interleukin-15 (IL-15) in lamina propria mononuclear cells (LPMNCs) and the expression of proliferation marker Ki67, FAS and TFR on epithelial cells (Maiuri et al. 2003a, Maiuri et al. 2000). Similar effects were reported for biopsy samples exposed to IL-15. Enterocyte apoptosis induced by gliadin and IL-15 could be abolished by anti-IL-15 antibodies. Recently, it was shown that IECs also produce increased levels of IL-15 after exposure to gliadin peptides (Barone et al. 2011). Interleukin-15 produced by dendritic cells and macrophages induces strong surface expression of MICA (Hue et al. 2004) a receptor for NKG2D expressed by cytotoxic CD8 αβ- and γδ-T cells. An interaction between MICA and NKG2D induces enterocyte apoptosis by intraepithelial lymphocytes (IELs) which results in extensive mucosal damage and the continued
introduction of gliadin peptides into the mucosa. Furthermore, IL-15 promotes the production of IFNγ by IELs (Di Sabatino et al. 2006, Garrote et al. 2008). Interleukin 15 may also act in an autocrine manner as it has been shown to induce anti-apoptotic pathways in IELs (Malamut et al. 2010).

In response to gliadin exposure, intestinal epithelial cells (IECs) have been shown to release zonulin (later identified as prehaptoglobin-2) in the cell medium (Drago et al. 2006, Tripathi et al. 2009). Zonulins binding to the cell surface are suggested to induce changes in cell cytoskeleton and increase monolayer permeability presenting a paracellular route into the mucosa. Apical to basal retrotranscytosis of gliadin peptides mediated by the transferrin receptor (TFR) CD71 and facilitated by secretory IgA (sIgA) antibodies has been proposed (Matysiak-Budnik et al. 2008, Heyman, Menard 2009). Apoptosis of enterocytes by gliadin peptides and direct sampling of gluten peptides in intestinal lumen by dendritic cells (DCs) are other proposed mechanisms (Chieppa et al. 2006, Rescigno et al. 2001). The transcellular pathway for gliadin transport has been associated with enterocyte proliferation and oxidative stress. In epithelial cells, the gliadin peptides are endocytosed which delays the maturation of early endosomes into late endosomes (Barone et al. 2010). These peptides were also shown to prolong the activation of the epidermal growth factor receptor (EGFR) pathway, which was implicated in the proliferation of epithelial cells (Barone et al. 2007b). It was recently shown that overnight stimulation of Caco-2 cells with gliadin peptides increases IL-15 gene transcription, which is dependent on preexisting IL-15 (Barone et al. 2011). Interestingly, the gliadin induced proliferation was found to be dependent on IL-15 and EGFR pathways. In other studies, effects like the disruption of epithelial tight junctions (TJs) (Ciccocioppo et al. 2006, Sander et al. 2005), cellular oxidative and endoplasmic reticulum (ER) stress (Luciani et al. 2009, Caputo et al. 2012) and alterations in actin cytoskeleton and protein trafficking (Reinke et al. 2011) have been reported in epithelial cells which are exposed to pepsin and trypsin digested gliadin (PT-G). It is, however, not certain which of these mechanisms is the primary method for gliadin translocation across intestinal epithelial layer. It is possible that gliadin peptides enter the mucosa through several different mechanisms and many are still unknown.
Figure 2. The innate and adaptive immune responses in celiac disease. sIgA: secretory immunoglobulin A class antibodies, CD71: transferrin receptor, p31-43 and p57-68: gliadin derived peptides corresponding to the amino acids 31-43 and 57-68, respectively. INFγ: interferon gamma, IL-15: interleukin-15, NK cell: natural killer cell, tTG: tissue transglutaminase; Q: glutamine, E: glutamate, IEL: intraepithelial lymphocyte, DC: dendritic cell.

Tissue transglutaminase (tTG) or transglutaminase 2 (TG2) belongs to a family of calcium-dependent enzymes that catalyze the crosslinking of proteins and participates in wound healing process. tTG is normally present in the cytoplasm but can also be released during an injury. It’s role as an autoantigen in CelD and gliadin as its preferred substrate was first described by Dieterich et al. (Dieterich et al. 1997). Both, CelD associated HLA-DQ2 and –DQ8 molecules (expressed by APCs) have a preference for binding peptides that have negatively charged residues at certain positions (Vartdal et al. 1996, Godkin et al. 1997). It was subsequently shown that tTG selectively deamidates glutamine residues in gliadin to negatively charged glutamic acid (also known as glutamate) (van de Wal et al. 1998), thus highly improving its affinity for DQ2 and DQ8 molecules on APCs. The adaptive T-cell response to gluten in CelD is mainly driven by CD4+ T (Th1) cells. The cells have a cytokine profile of Th1 cells (Nilsen et al. 1995, Nilsen et al. 1998), which is dominated by IFNγ. Tregs are also present in the intestine and play an important role in the induction of tolerance to self-antigens and dietary proteins (Gianfrani et al. 2006). However, several studies report impairment in the function of Treg cells (Granzotto et al. 2009, Hmida et al. 2012) in CelD. Following the discovery of Th17 cells and their involvement in the pathogenesis of other autoimmune diseases, it was speculated that
Th17 cells may also be responsible for driving mucosal inflammation in CelD. This idea was particularly interesting as an impairment of T\(_{reg}\) cell function had been reported. Results from studies on Th17 cells in CelD have been rather contradictory. Monteleone et al. found higher expression of IL-17A RNA and protein in active CelD biopsy samples compared to inactive CelD and normal mucosal biopsy samples (Monteleone et al. 2010). The authors suggested that in active CelD patients, IL-17A was mainly produced by CD4+ T cells and double positive CD4-CD8+ T cells. Similarly, Castellanon-Rubio et al. found signatures of both Th1 and Th17 CD4+ T cells in active CelD (Castellanos-Rubio et al. 2009). In contrast to these studies, Bodd et al. reported that gluten-specific T cells produce IFN\(\gamma\) and IL-21 but not IL-17 or IL-22 (Bodd et al. 2010). Thus, the role of Th17 cells in CelD is elusive. Nonetheless IL-21 may have a bigger role than hitherto anticipated as it targets, among others, CD8+ T cells, NK cells and is important for B cell differentiation and plasma cell function.

An autoimmune reaction to a self-protein, i.e. tissue transglutaminase (tTG), is mounted in CelD individuals. Immunoglobulin-A anti-gliadin antibodies (AGA) (Unsworth et al. 1981), IgA antibodies targeting connective tissue fibers, i.e. anti-reticulin antibodies (ARA) (Lerner, Kumar & Iancu 1994) and those targeting structures surrounding smooth muscle fibers, i.e. anti-endomysium autoantibodies (EMA) (Burgin-Wolff et al. 1991) and IgA (also IgG) anti-tissue transglutaminase antibodies (ATGA) are produced by plasma cells in the mucosa of CelD patients (Figure 2). In the small intestinal mucosa of CelD patients, these autoantibodies are found as IgA deposits below the basement membrane along the villous and crypt epithelium and around the blood vessels (Korponay-Szabo et al. 2004, Tosco et al. 2008). In active CelD, celiac-specific IgA antibodies are also found in the serum. Interestingly, these autoantibodies are gluten dependent and disappear from the system upon gluten withdrawal in the diet (Sulkanen et al. 1998). With the identification of tTG as the CelD autoantigen (Dieterich et al. 1997) it soon became clear that both, EMA and ARA, target tTG (Korponay-Szabo et al. 2000). The mechanism of production and significance of autoantibodies against tTG in CelD pathogenesis are not clear. It is conceivable that tTG-specific B cells endocytose tTG-gliadin complexes and process them intracellularly (Sollid et al. 1997). The resulting gluten peptides are then presented to gluten-specific CD4+ T cells which activate B cells to produce antibodies against tTG. The tTG autoantibodies have been shown to promote epithelial cell proliferation (Barone et al. 2007a) and inhibit differentiation (Halttunen, Maki 1999), facilitate the passage of gliadin peptides across the epithelial layer (Rauhavirta et al. 2011) and promote the maturation of dendritic cells and their ability to stimulate T cells in a concentration dependent manner (Dalleywater, Chau & Ghaemmaghanii 2012).
2.5. Diagnosis and treatment

In May 1989, the European Society of Pediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) organized a workshop on “Diagnostic criteria of celiac disease” in Budapest, Hungary. The outcome of that meeting was presented in 1990. The ESPGHAN proposed a revised protocol for Celiac Disease (CelD) diagnosis (Walker-Smith et al. 1990). Following new findings in the field of CelD, the criteria for the diagnosis has been revised again after a period of 20 years (Husby et al. 2012). The new report concludes that the diagnosis of CelD depends on symptoms which are dependent on gluten, the levels of CelD-specific autoantibodies, the presence of HLA-DQ2 and/or HLA-DQ8 and changes in the duodenal biopsy characteristic of histological changes in CelD (i.e. villous atrophy and crypt hyperplasia). A clinical remission and relief from all symptoms on a strict GFD should suffice for diagnosis with CelD. In asymptomatic patients, however, a gluten challenge and multiple control biopsies may be required to ascertain CelD.

The examination of a small intestinal biopsy for mucosal changes (villous atrophy and crypt cell hyperplasia) and intraepithelial lymphocyte counts (≥25 per 100 enterocytes) is still the gold standard for CelD diagnosis. It is recommended that serological tests should be performed before referring a patient for intestinal biopsy and commencement of a GFD (Walker, Murray 2011). The contemporary approach for CelD diagnosis is to check serum levels of IgA (or IgG where relevant) ATGA, EMA or antibodies against deamidated gliadin peptides (Gujral, Freeman & Thomson 2012, Niveloni et al. 2007). Due to low sensitivity and specificity, the AGAs are very rarely used in diagnosis. The EMA and ATGA, on the other hand, show high levels of specificity and sensitivity (Walker, Murray 2011, Rostom et al. 2005). Despite high sensitivity and specificity (>91% and 99% respectively), a low positive predictive value (29%) has been demonstrated for tTG autoantibodies by Hopper et al. (Hopper et al. 2008). In the same study, when IgA tTG testing was combined with IgA EMA serological test, the positive predictive value was found to be 71.7%. Thus, using a two-step approach is also recommended. Serology is likely to miss some cases as 10-20% of untreated CelD patients may remain negative for serum EMA (Dickey, Hughes & McMillan 2000b) and may also exhibit less prominent small bowel histology (Abrams et al. 2004). Patients may also be tested for HLA-DQ2 or DQ8 molecules as >97% of CelD patients carry these molecules (Walker, Murray 2011) but a positive test does not confirm CelD as genes encoding these molecules are quite common (25-40%) in the Western populations. A negative test is, nonetheless, a strong negative predictor of CelD. It is recommended that in the presence of negative serology, if the clinician remains suspicious of CelD, a biopsy examination should be undertaken. It has also been shown that in EMA-negative patients, gluten-dependent mucosal IgA tTG deposits may be present (Salmi et al. 2006). The
mucosal IgA deposits, in the absence of villous atrophy, are a good indicator of forthcoming CelD (Salmi et al. 2006, Kaukinen et al. 2005). Moreover, mucosal IgA tTG deposits are also useful in the follow up of patients with regards to small bowel mucosal recovery as they disappear slowly on a GFD compared to serum IgA levels (Koskinen et al. 2010, Dickey, Hughes & McMillan 2000a).

A strict GFD is currently the only treatment for CelD. A life-long adherence is undoubtedly a daunting undertaking. However, the challenges associated with a GFD are outnumbered by health benefits. A number of alternative therapies/approaches (such as modification of toxic gliadin peptides, prevention of toxic gliadin peptide absorption, gliadin deamidation blockade and development of vaccines among others) are currently being developed (Gujral, Freeman & Thomson 2012), including detoxification of gluten peptides (Stenman et al. 2009, Stenman et al. 2010) and methods to prevent their uptake in the intestine.

2.6. Celiac disease and associated disorders

Individuals suffering from CelD have an increased risk for developing other diseases, many of which are of autoimmune nature. Collin et al. investigated the incidence of various CelD associated diseases in a group of 335 CelD patients and 335 age and sex matched control subjects with gastrointestinal symptoms, other than CelD (Collin et al. 1994). They reported that compared to the control group, CelD patients had a higher incidence of endocrine disorders (4.3% vs. 11.9%), T1D (1.5% vs. 5.4%), connective tissue diseases (2.7% vs. 7.2%), Sjögren’s syndrome (0.3% vs. 3.3%) and autoimmune thyroid diseases (2.7% vs. 3.6%). Subsequently, in a reciprocal manner, increased CelD incidence has been reported in T1D patients (Bhadada et al. 2011, Arato et al. 2003). In a study by Book et al., ten out of 97 Down’s syndrom patients had CelD (Book et al. 2001). In a group of 140 pediatric patients suffering from autoimmune liver disease, 23 patients also had CelD (Caprai et al. 2008). In this study, all 23 CelD patients remitted when they were put on a GFD in combination with immunosuppressive therapy. However, 14 patients had a relapse because they discontinued the immunosuppressive therapy or resumed gluten containing diet. Recently, in a Swedish study comprising 29000 patients with biopsy-proven CelD, a three-fold higher risk for SLE in CelD patients compared to the general population was reported (Ludvigsson et al. 2012). Why CelD patients, or autoimmune disease-affected individuals in general, are more prone to developing other autoimmune diseases is not clear. An explanation may lie in the fact that HLA risk haplotypes and many other risk genes are shared among different immune-mediated diseases. It is possible that different combinations of genes may be responsible for different disease outcomes.
Dermatitis herpetiformis (DH) is an itchy and blistering skin disease, which predominantly affects the exterior surfaces of elbows, knees, buttocks, back and scalp. DH is an extra-intestinal manifestation of CeID, which appears in about 25% of CeID patients who may or may not have gastrointestinal symptoms (Collin, Reunala 2003). Although it was first described in 1884 by Louis Duhring (Duhring 1983), its association with CeID was reported in 1966. (Marks, Shuster & Watson 1966) It is rather common in individuals with Northern European ancestry and affects more men than women with a male-to-female ratio ranging from 1.5:1 to 2:1. (Gawkrodger et al. 1984, Smith et al. 1992) Although it is considered relatively uncommon, several cases have been reported in Asian countries (Ohata et al. 2012, Sawhney, Singh 2011, Singal, Bhattacharya & Baruah 2002, Zhang et al. 2012). Earlier studies have reported the prevalence of DH to range from 10.4 to 66 per 100 000 individuals and the incidence rate has ranged from 0.4 to 2.6 per 100 000 individuals per year. (Gawkrodger et al. 1984, Reunala, Lokki 1978, Moi 1984, Collin et al. 1997) However, a recent study reported an increased prevalence (75.3 per 100 000 individuals) and incidence (3.5 per 100 000 individuals) of DH in Finland (Salmi et al. 2011). The age at DH onset is typically 30-40 years but the disease can also appear in the childhood or in older ages. (Reunala 2001, Ermacora et al. 1986) In addition, the first degree relatives of affected individuals are at greater risk for developing CeID and/or DH. In a study by Hervonen et al. the overall risk in first degree relatives was estimated to be 15 times higher than in the general population with the siblings being the most disease prone group among them (Hervonen et al. 2002). High concordance rates have been reported among siblings and twin pairs, where both twins are affected with CeID and DH or with either of the two diseases (Hervonen et al. 2000, Karell et al. 2002).

Despite its extraintestinal mode of presentation, DH is strongly associated with the HLA class II region on chromosome six. Spurkland et al. conducted a study including 50 DH patients and 290 healthy control individuals (Spurkland et al. 1997). They found that genes encoding the HLA-DQ2 heterodimer were present in 86% of the patients and 25% of the controls and another 12% of the cases carried genes encoding the HLA-DQ8 heterodimer. Thus HLA-DQ2 or HLA-DQ8 positivity is a major genetic risk factor for DH in Caucasians. Genes encoding these heterodimers are present in other populations also, albeit at a lower frequency (Liu, Cheng 2007). Recently, 91 DH patients of Japanese descent were found to have HLA genotypes distinct from HLA-DQ2 and –DQ8 (Ohata et al. 2012), implying that population-specific genetic risk factors are likely to exist. In addition to the strongly associated HLA region, rare genetic variants of TNF family of genes were found to be present at a higher frequency in DH patients than in healthy controls (Messer et al. 1994).
Furthermore, in a study comprising CelD patients with DH, linkage to the 11q and 5q chromosomal regions, which harbor potentially relevant candidate genes for immunological diseases, was suggested (Holopainen et al. 2001). Another study by Koskinen et al. also provides some evidence of DH linkage to the 19p13 chromosomal region (Koskinen et al. 2008). The 19p13 region harbors a myosin gene, *Myosin IXB (Myo9B)*, which encodes a protein with actin remodeling properties and a potential role in the maintenance of intestinal barrier. Independent linkage to and association with CelD of this region has been shown previously (Van Belzen et al. 2003, Monsuur et al. 2005).

2.7.1. Diagnosis and treatment of dermatitis herpetiformis

DH diagnosis is made based on the presence of granular IgA deposits in the dermal papillae demonstrated by direct immunofluorescence (DIF) examination (van der Meer 1969). However, the presence of AGA and ARA and EMA has also been demonstrated in DH patients (Vainio et al. 1985, Hallstrom, Reunala 1985, McCord, Hall 1994). In a study by Beutner et al., DH cases were found to have negative DIF findings for junctional IgA deposits in two biopsy specimens but positive for IgA EMA and ATGA (Beutner et al. 2000). The presence of ATGA and EMA in DH has been suggested by other studies as well (Kumar et al. 2001, Chorzelski et al. 1983, Volta et al. 1992). As the use of DIF and serum tests for EMA and ATGA in DH cases not detected by DIF alone, the use of both tests in parallel for DH screening has been suggested (Kumar et al. 2001). Nevertheless, the test for EMA is done using an indirect immunofluorescence (IIF) method on sections of monkey esophagus (Chorzelski et al. 1983) making it both time-consuming and subjective. A serum test for the detection of ATGA has also been proposed as an alternative to DIF for DH diagnosis (Desai, Krishnan & Hsu 2005). ATGA are also found in CelD patients without DH, rendering this method less effective. As in CelD, tTG was thought to be the major autoantigen in DH. It was later shown that the epidermal transglutaminase (eTG) is the autoantigen in DH against which the IgA antibodies are directed (Sardy et al. 2002). This observation finds support in recent studies where IgA antibodies against eTG correlated more strongly with DH (treated or untreated) than with CelD (Rose et al. 2009, Borroni et al. 2012). The authors concluded that IgA antibodies against eTG are the most sensitive serological marker in DH. Furthermore, as in CelD, a GFD is the only effective treatment for DH and is recommended even in the absence of gastrointestinal symptoms (Collin, Reunala 2003). DH patients are often advised to complement GFD with dapsone, a drug frequently used to soothe the rash symptoms (Reunala 2001). Strict adherence to a GFD usually mitigates the GI symptoms within a few weeks, however relief from rash may take a few months (Collin, Reunala 2003). A good response to a GFD has been reported in both pediatric and adult DH (Ermacora et al. 1986, Garioch et al. 1994).
Although a GFD is highly effective in soothing gluten-triggered enteropathy, non- or partial-compliance with a GFD may lead to the development of other diseases.

2.7.2. Dermatitis herpetiformis and associated diseases

A higher incidence of several immune-mediated or other malignancies among DH patients has been reported. Reunala et al. studied a cohort of 305 DH patients over a period of ten years and found that different diseases were associated with DH (Reunala, Collin 1997). They reported that 4.3% of the DH patients had developed autoimmune thyroid disease, 1% had T1D, 1.3% had SLE, 1% had Sjögren’s syndrome, 1.3% had sarcoidosis and 1.6% of the DH patients had developed vitiligo or alopecia areata. RA, UC, SLE and splenomegaly have also been reported to have a high incidence among DH patients (Davies, Marks & Nuki 1978, Kurano, Lum & Izumi 2010). In a study comprising 50 DH patients, 26 were found to have thyroid abnormalities (Cunningham, Zone 1985) whereas Hervonen et al. reported an increased prevalence of T1D in DH patients (2.3%) and their first-degree relatives (3.0%) compared with controls (Hervonen et al. 2004). High incidences of dental-enamel defects and chronic gastritis have also been reported in DH patients (Alakoski et al. 2012, Aine, Maki & Reunala 1992). Perhaps the most studied of associated diseases with DH is the association between DH and lymphoma. A majority of the lymphomas have been classified as non-Hodgkin’s lymphoma, however enteropathy-associated T-cell lymphoma (EATL) and B-cell lymphoma have also been reported (Sigurgeirsson, Agnarsson & Lindelof 1994, Hervonen et al. 2005, Viljamaa et al. 2006). A significantly higher risk for non-Hodgkin’s lymphoma (relative risk of 5.4) in male patients was reported in a study conducted on Swedish cancer patients with DH (Sigurgeirsson, Agnarsson & Lindelof 1994). Furthermore, a protective effect of GFD on the development of lymphoma in DH patients has been demonstrated. Multiple studies have shown that only those DH patients who did not comply with a GFD or had been on a GFD for a period of less than five years had developed malignancies (Viljamaa et al. 2006, Leonard et al. 1983, Lewis et al. 1996).

3. Inflammatory bowel disease

Inflammatory bowel disease (IBD) is a remitting and relapsing chronic inflammation of the GI tract, precipitated by unknown environmental factors in individuals with a certain genetic background. The two major forms of IBD are the closely related inflammatory disorders Crohn’s disease (CD) and ulcerative colitis (UC). In about 10% of cases, it is not possible to differentiate between CD and UC and the term IBD unclassified (IBDU) may be used. In UC, the inflammation is mainly restricted to the colonic mucosa, whereas in CD it can affect any part of the GI tract. Both UC and CD share several characteristic features but can be distinguished from each other on the basis of predisposing genes,
environmental risk factors, clinical, endoscopic and histological features. These features are listed in Table 3. UC and CD have been shown to share a number of susceptibility genes (Anderson et al. 2011, Franke et al. 2010). A wealth of information is now available on various risk factors, genetic and environmental, but no cure is yet available. The condition usually worsens with time and many patients have to undergo surgical procedures and take immunosuppressive drugs.

Table 3. Features to Differentiate UC from CD

<table>
<thead>
<tr>
<th>Clinical features</th>
<th>Crohn’s disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diarrhea with urgency, often small volumes but frequent</td>
<td>Diarrhea (often) accompanied by abdominal pain and malnutrition</td>
</tr>
<tr>
<td>Bloody diarrhea is predominant</td>
<td>Stomatitis (inflammation of the mouth)</td>
</tr>
<tr>
<td>Endoscopic and radiological</td>
<td>Perianal lesions</td>
</tr>
<tr>
<td>Diffuse superficial colonic inflammation</td>
<td>Discontinuous transmural asymmetric lesions</td>
</tr>
<tr>
<td>Rectum is involved, inflammation may be patchy</td>
<td>Mainly involving ileum and right- side colon</td>
</tr>
<tr>
<td>Shallow erosions and ulcers</td>
<td>Cobblestone appearance</td>
</tr>
<tr>
<td>Spontaneous bleeding</td>
<td>Longitudinal ulcers, Aphthous ulceration</td>
</tr>
<tr>
<td>Histopathological</td>
<td>Deep fissures</td>
</tr>
<tr>
<td>Diffuse inflammation in mucosa or submucosa</td>
<td>Granulomatous inflammation</td>
</tr>
<tr>
<td>Crypt architecture distortion</td>
<td>Fissures or aphthous ulcers can be seen; often transmural inflammation</td>
</tr>
</tbody>
</table>

3.1. Crohn’s disease (CD)

CD is a systemic inflammatory disease with episodes of remission and relapse. It is predominantly a disease of the gut which may also have extraintestinal manifestations. CD can affect the entire GI tract, from mouth to anus. Typically, CD patients suffer from a discontinuous inflammation in various parts of the GI tract with the development of complications like strictures, abscesses and fistulas. CD patients are phenotyped according to the Montreal classification of IBD (Silverberg et al. 2005). This classification takes into account the age (A) of disease onset, location (L) of the disease and disease behaviour (B) as the predominant phenotypic elements. The age has been further divided into three subgroups (A1: < 16 years, A2: 17-40 years and A3: > 40 years). The location (L)-group comprises four subgroups (L1: ileal, L2: colonic, L3: ileocolonic and L4: upper GI tract). The L4 location acts as a modifier that can be added to L1-L3 when concomitant upper GI disease is present. The disease
behavior is defined as B1 (non-stricturing, non-penetrating), B2 (stricturing) and B3 (penetrating). In addition, the perianal disease modifier (p) is added when concomitant perianal disease is present. All these features are shown in Figure 3.

CD is dynamic in nature as its behavior has been reported to change over time. In a population-based study, 306 patients were diagnosed between the ages (A2 category) of 17 and 40 years and the proportion of patients with disease location in L1, L2, L3 and L4 was 45%, 32%, 19% and 4% respectively (Thia et al. 2010). Furthermore, 81%, 5% and 14% of the patients had a B1, B2 and B3 or B3p phenotype, respectively. Altogether 19% of these patients progressed to a more aggressive phenotype within 90 days after diagnosis and over half of the patients (51%) experienced an intestinal complication within 20 years after diagnosis. In another study, Louis et al. reported that, 10 years after initial diagnosis, 45.9% of patients had a change in disease behavior (Louis et al. 2001). The most prominent change was from non-stricturing, non-penetrating phenotype to either stricturing (27.1%) or penetrating (29.4%) phenotype. While age at diagnosis had no impact on this, ileal CD was more often stricturing and colonic or ileocolonic CD was more often penetrating. Studies in different cohorts indicate that one year after diagnosis, 10-30% of CD patients experience disease aggravation, 15-25% have low activity and 55-65% are in remission. (Munkholm et al. 1995, Loftus, Schoenfeld & Sandborn 2002). On long-term, 13-20% of CD patients have a chronic active course of disease activity, while 67-73% have a chronic intermittent course. Only a small proportion (10-13%) of CD patients remains in remission for several years. In a study by Jess et al., increased mortality compared to the general population was reported in CD patients (Jess et al. 2002). The CD-related mortality was particularly higher in women compared to men.

3.2. Ulcerative colitis

UC is a relapsing inflammatory disease which involves the rectum and colon and extends in a continuous retrograde mode. The inflammation is generally diffuse and superficial. Depending on the anatomic location of inflammation, patients can be referred to as having ulcerative proctitis (involvement limited to the rectum), left-sided colitis (involving the sigmoid colon with or without the descending colon) or extensive colitis (also known as pancolitis) (Silverberg et al. 2005). The typical features of UC are shown in Figure 4. In patients with pancolitis, the last few centimeters of the ileum can also be inflamed. Such a backwash ileitis remains superficial and does not show deep ulcerations, stricture or fistula. Clinically, disease activity is described as mild (up to four bloody
Figure 3. The phenotypes of CD according to the Montreal classification. Modified from Baumgart and Sandborn (2012) and reprinted with permission.

stools daily and no systemic toxicity), moderate (four to six bloody stools daily and minimal toxicity), severe (more than six stools daily and signs of toxicity such as fever anemia, raised erythrocyte sedimentation rate) or fulminant (Carter et al. 2004, Kornbluth, Sachar & Practice Parameters Committee of the American College of Gastroenterology 2004). The patients with fulminant disease may have more than ten stools daily, continuous bleeding, toxicity, abdominal tenderness and distension. They may also require blood transfusion and colonic dilation is visible on abdominal plain films. The course of the disease changes over time. In a study by Langholz et al., comprising 1161 UC patients, 9.1% of the cases had fulminant UC, 70.7% had moderate to high disease activity and 20.2% had low disease activity (Langholz et al. 1991). In 18% of the UC patients, the whole colon was
affected (pancolitis), 36% had left-sided colitis and 44% had UC proctitis. In another study comprising 515 UC patients, evaluation of the disease spread over-time revealed that 39% had progressed to a more advanced stage (Langholz et al. 1996). Of these, 4% had developed extensive colitis and in 35%, the inflammation had spread to involve substantial part of the large bowel. Patients with extensive colitis have higher a risk for colectomy than those with proctitis (Hoie et al. 2007b). Despite severe manifestations, UC patients do not have higher mortality rates than the general population (Hoie et al. 2007a).

3.3. Clinical features of IBD

In general, symptoms in IBD depend upon the GI segment which is affected. The symptoms may be mild or severe during relapses or they may decrease or completely disappear while the patient is in remission (Bernstein et al. 2010). IBD patients may present with symptoms like diarrhea (stool may contain mucus or blood, nocturnal diarrhea or incontinence), constipation (can be the primary symptom of UC which is limited to rectum), pain or rectal bleeding with bowel movement, severe bowel movement urgency, tenesmus (straining to urinate or defecate, without the ability to do so), abdominal cramps and pain (in CD in the lower right quadrant or around the umbilicus; in moderate

Figure 4. Typical features of UC are shown in the figure.
to severe UC in the lower left quadrant) and nausea and vomiting (more common in CD than in UC). Other symptoms which are more often associated with CD than with UC are fever, loss of appetite, weight loss, fatigue, night sweats, growth retardation and primary amenorrhea.

Pediatric IBD patients present with the same symptoms as adult IBD patients but tend to have more extensive disease than adults at diagnosis (Langholz et al. 1997). Similarly, CD patients less than five years of age tend to have significant large bowel involvement (Kelsen, Baldassano 2008). This group may have more significant diarrhea (91%) at presentation compared to older children and adults (65%). Other symptoms which are more common in young children with CD are growth failure, anorexia, malaise, jaundice, late-onset puberty and osteopenia/osteoporosis (Kim, Ferry 2004). A retrospective pediatric cohort study of CD, which included 566 boys and 423 girls (age at diagnosis 0-17 years), reported that female CD patients had overall more severe course of disease but boys were at a greater risk for growth failure (Gupta et al. 2007).

Up to 25% of IBD patients suffer from extraintestinal manifestations (EIM) of IBD and include arthralgia (pain in joint), arthritis (acute or chronic inflammation of a joint), ankylosing spondylitis (a chronic inflammatory disease of the axial skeleton), pyoderma gangrenosum (a chronic skin disease characterized by large spreading ulcers), erythema nodosum (an inflammation of the fat cells under the skin), iritis (inflammation of the iris of the eye), uveitis (inflammation of the uvea) and episcleritis (inflammation of the episcleral tissue) (Bernstein et al. 2010). Osteoporosis, venous thromboembolism (a blood clot that forms within a vein), avascular necrosis (cellular death of bone components) and mood disorders such as anxiety and depression are increased in IBD compared to the general population. Many of these complications may exist concomitantly in IBD patients and some may run an independent course from the IBD. In a study comprising 1649 pediatric IBD patients (28% of the patients were diagnosed with IBD at age less than six years) six percent had at least one EIM before diagnosis of IBD (Jose et al. 2009). The prevalence of EIM among children diagnosed before the age of six years and among older children was 3% and 7%, respectively. Arthritis (26%) and aphthous stomatitis (21%) were the most common before IBD diagnosis, whereas arthritis (17%) and osteopenia/osteoporosis (15%) were the most common EIM after IBD diagnosis.

Intestinal complications are common in IBD (Bernstein et al. 2010). Profuse bleeding (hemorrhage) from ulcers is more common in UC than in CD. However, massive bleeding in CD is more often seen from ileal ulceration than colitis. Bowel perforation and intra-abdominal abscesses (a localized collection of pus often accompanied by swelling and inflammation) may also occur in CD (Poritz, Koltun 2007). Strictures in CD are often inflammatory and possible to resolve with medical treatment, whereas colonic strictures in UC are presumed to be malignant until proven otherwise.
Fistulas and perianal disease are the hallmarks of CD and surgical intervention is required in cases which do not respond to vigorous medical treatment or where abscesses have developed. Toxic megacolon (Hokama et al. 2012, Strong 2010) is a rare life-threatening colitis which is more common in UC than in CD. A surgical intervention is necessary if this condition is not resolved by medication within 24 hours.

3.4. Epidemiology of IBD

Inflammatory bowel disease may affect an individual at any age, but it generally starts in young adulthood and differs in the age of onset for UC and CD. The peak age for UC onset is usually 30-40 years, whereas for CD it is 20-30 years (Cosnes et al. 2011). However, in 15-25% of cases IBD appears before 20 years of age (Kim, Ferry 2004, Oliva-Hemker, Fiocchi 2002). Pediatric IBD accounts for about 7-20% of all IBD cases and a majority of the cases are diagnosed between ten and 14 years of age. However, a small proportion (4%) of pediatric IBD is diagnosed before the age of five. A study which covered 56.5% of Swedish pediatric population reported that the median age at diagnosis for UC, CD and IC was 12.2 years, 13 years and 11.2 years respectively and 7.5% of the children were five years of age or less (Lindberg et al. 2000). In a study by Turunen et al. on a Finnish pediatric cohort, the mean age at diagnosis for UC, CD and IBDU was found to be 10.9 ± 4.1 years, 11.9 ± 3.7 years and 10.8 ± 4.8 years, respectively (Turunen et al. 2006). Fourteen percent of the children were reported to be less than six years of age at the time of diagnosis in this study. This seems to be the general picture in the onset of pediatric IBD as similar results have been reported in other parts of the world (Gupta et al. 2007, Treepongkaruna et al. 2006). The incidence of UC is higher in men, whereas that of CD is higher in women (Bernstein et al. 2006, Molodecky et al. 2012). However, several studies conducted in Asia and the Middle East report higher incidence of CD in the male population (Molodecky et al. 2012). In a recent report by the World Gastroenterology Organization, it was asserted that despite the higher CD prevalence among female pediatric patients, the CD incidence rates have been greater among males over the past few years (Bernstein et al. 2010). Thus, an equalization of sex distribution may occur.

The incidence and prevalence of IBD has increased over time, with more cases per year being reported in Southern and Eastern Europe, Asia and in other developing countries which were previously regarded as low-incidence areas (Ng et al. 2013). In a recent systematic literature review, the incidence of UC (person-year per 100 000 individuals) in Europe, Asia and the Middle East and in North America was reported to be 24.3, 6.3 and 19.2 respectively (Molodecky et al. 2012). The corresponding values for CD were 12.7, 5.0 and 20.2 respectively. The prevalence of UC (person-year per 100 000 individuals) in Europe ranged from 4.9 to 505, from 4.9 to 168.3 in Asia and the Middle
East and from 37.5 to 248.6 in North America while the CD prevalence rates were 0.6-322, 0.88-67.9 and 16.7-318.5 for Europe, Asia and the Middle East and North America, respectively. In Europe, the highest incidence rate for UC was reported in Iceland (24.3 per 100 000) and for CD in the United Kingdom (10.6 per 100 000) (Thompson et al. 1998). However, in a recent study by Jussila et al. for the study period 2000-2007, the mean annual incidence of IBD, UC and CD (per 100 000) was reported to be 34, 24.8 and 9.2 respectively (Jussila et al. 2012). In North America, Canada has the highest incidence rates for UC and CD at 19.5 and 20.2 per 100 000 respectively (Bernstein et al. 2006). The incidence of IBD has been higher in developed countries and has long been associated with a westernized lifestyle. In the Pacific, Australia and New Zealand are highly developed and industrialized nations, which also share genetic background with Europeans and North Americans. Not surprisingly, high IBD incidence rates have been reported at these places (Gearry et al. 2010, Wilson et al. 2010).

A similar increase in the incidence and prevalence of IBD in children has been observed. In Sweden, an increase in the incidence of pediatric UC (per 100 000) was reported from 4.6 in 1984-1986 to 7.0 in 1993-1995 (Lindberg et al. 2000). No change in the incidence of CD and IBDU was however reported. In Finland, the pediatric IBD incidence has increased three-fold (from 5 to 15) from 1987 to 2003 with UC being the most prevalent disease (Lehtinen et al. 2011). During the years 1996-2006, the annual incidence of pediatric UC per 100 000 increased from 1.8 to 4.9 in Northern California (Abramson et al. 2010). In contrast to high UC incidence rate observed in these studies, an increase in the incidence of CD in children (aged 0-19 years) from 3.4 per 100 000 in 1988-1990 to 5.9 per 100 000 in 2006-2007 was reported in France (Chouraki et al. 2011). A similar trend has been reported from other studies in the United Kingdom (Cosgrove, Al-Atia & Jenkins 1996, Henderson et al. 2012). What affects the differential rise in the incidence of UC and CD in different regions is not clear, but both genetic and environmental factors are likely to be responsible for this trend.

IBD is more prevalent in developed and industrialized countries compared to the developing countries. In developed countries an increased incidence of IBD has been reported in urban areas compared to rural areas (Ekbom et al. 1991, Soon et al. 2012). The increase in IBD incidence in urban areas may be linked to dietary habits as multiple studies demonstrate that high animal protein intake, high sugar consumption and fat-rich food are risk factors for IBD (Sakamoto et al. 2005, Russel et al. 1998a). A rise in the occurrence of IBD has been observed in countries, which have adopted a westernized lifestyle, including India, China, Iran and Korea (Zheng et al. 2010, Desai, Gupte 2005, Firouzi et al. 2006, Yang et al. 2008). The exact factors contributing to an elevated risk are not known but may include, among others, changes in diet, improved sanitation and hygiene...
standards. Indeed, in Eastern European countries such as Hungary and Croatia, IBD incidence rates comparable to those in Northern Europe have been reported (Lakatos et al. 2011, Sincic et al. 2006). In contrast, epidemiological studies from Romania and Poland report continued low-incidence rates for IBD (Gheorghe et al. 2004, Wiercinska-Drapalo et al. 2005). Whether the difference in incidence rates among these countries is real or due to differential diagnostic criteria is unclear.

It has been proposed that higher exposure to sunlight and consequently higher vitamin D levels may have a protective effect on IBD progression. A study performed on 175,000 female American nurses found that increasing latitude of residence was associated with higher risk for IBD (Khalili et al. 2012). Although this may hold true for North America and Northern European countries, it does not hold true for India where sunlight is found in abundance for most part of the year (Molodecky et al. 2012, Desai, Gupte 2005). Smoking is associated with increased risk for CD (Lindberg, Jarnerot & Huitfeldt 1992, Russel et al. 1998b) but with protection from UC (Odes et al. 2001, Higuchi et al. 2012). Despite positive associations reported between smoking and CD, a lack of association between the two has also been reported (Carlens et al. 2010, Reif et al. 1995, Reif et al. 2000). In a study comprising 339 affected sibling pairs who have similar genetic susceptibility for IBD (Bridger et al. 2002), smoking was associated with CD development (OR: 3.55) but protected from UC (OR: 0.28) and in 23 disease-discordant siblings pairs, the smoker had CD and the non-smoker had UC. Among other risk factors for IBD are appendectomy, repeated use of antibiotics in the first year of life and the use of oral contraceptives (Lopez-Serrano et al. 2010, Andersson et al. 2003, Virta et al. 2012, Cornish et al. 2008). The mechanistic details of how these factors affect the risk for IBD have not been outlined yet.

3.5. Genetics of IBD

Chronic inflammation in IBD results in severe mucosal damage. Although a dysregulated immune system is responsible for the inflicted mucosal injuries, it is the genetic architecture of the affected individual that lies at the heart of IBD pathogenesis.

3.5.1. Family studies in IBD

Familial aggregation has been reported in IBD. In a study published in 1963, Kirsner et al. outlined several studies where multiple members within a family were suffering from UC (Kirsner, Spencer 1963). The first report of familial UC was published already in 1934 where two brothers and a sister were concordant for UC. The authors of this study concluded that UC is more common in the families of probands with UC. Subsequently, Mayberry et al. reported that nine percent of CD patients had first-degree relative with CD or UC (Mayberry, Rhodes & Newcombe 1980). The risk of
siblings developing the disease was estimated to be 30-times greater than that of the general population. In another study, the prevalence of IBD among relatives was reported to be 7.9% and the other affected individual was more often a sibling (Monsen et al. 1987). The prevalence of UC was 15-times higher in relatives than in non-relatives and the prevalence of CD in first-degree relative of UC patients was almost 3.5-times higher than in non-relatives. One of the first population-based studies was performed on 637 patients with CD or UC in Copenhagen, Denmark (Orholm et al. 1991). The results from the study revealed that the first-degree relatives of CD patients had a ten-fold risk for CD compared to the general population. Similarly, a first-degree relative of UC patient had an eight-fold risk for UC. Furthermore, for a relative of a CD patient to get UC and for a relative of a UC patient to get CD, the relative risk was four-fold and two-fold respectively compared to the relatives of non-IBD individuals. These observations suggested that both CD and UC are genetically related and the genetic factor may be stronger in CD than in UC. This was further demonstrated in a UK study by Satsangi et al. In this study, of 317 CD and 825 UC patients, 41 (13%) and 90 (12%) patients respectively had a positive family history (Satsangi et al. 1994). Although these numbers are comparable, in 6.6% of the CD cases, the affected family member was a sibling compared to 3.75% in the UC cases. This suggests that first-degree relatives of CD patients are at higher risk of IBD than that of UC patients. As reported by Peeters et al., 132 of 640 probands with CD had a positive IBD family history compared to only 17 (of total 797) control individuals (Peeters et al. 1996). In addition, the risk was particularly high (10.4% after age adjustment) for the offspring of a CD affected individual and daughters were at higher risk (12.6%) of IBD compared to sons (7.9%). Similarly, the female first-degree relatives of CD probands had a higher risk of IBD than their male counterparts. As there is a high degree of gene sharing between the affected individual and the first-degree relatives, these studies strongly suggested a pivotal role for genes in the etiology of IBD.

3.5.2. Twin studies

Twin studies are a powerful tool, often used in complex diseases, to better understand the contribution of genes and environmental factors in disease etiology. The twin studies are useful for determining if something is genetic, and making estimates for how heritable the disease is. Monozygotic twin pairs are genetically very similar, but not 100% identical because of spontaneous mutations or epigenetic modifications in the womb or after birth (Fraga et al. 2005). Dizygotic twins, on the other hand, are genetically equivalent to siblings but born at the same time and share only 50% of their genetic material. Thus, in a monogenic disease with a very high penetrance, a close to 100% concordance rate would be expected between monozygotic twins and a 50% concordance rate between dizygotic twins. However, if a disease is caused by environmental factors, both identical
and non-identical twin pairs would have the same concordance rates. The reported concordance rates for monozygotic twins and dizygotic twins are different and also differ among different studies. The first twin study on 80 unselected, same sex twins with IBD was performed in Sweden (Tysk et al. 1988) and a concordance rate of 6.3% and 44.4% was reported for UC and CD respectively for identical twins. The corresponding numbers for non-identical twins were zero percent and 3.9%, respectively. In other studies, the monozygotic concordance rates for CD ranges from 20% to 55% and those for dizygotic twins are between zero and 6.5% (Orholm et al. 1991, Tysk et al. 1988, Halfvarson 2011, Halfvarson et al. 2006, Halfvarson et al. 2003, Thompson et al. 1996, Spehlmann et al. 2008). The concordance rates for UC for monozygotic and dizygotic twin pairs have been reported to range from 6.3% to 16-7% and from 0.0% to 6.3%, respectively. However, similar to earlier population-based studies, twin studies also suggest that genetic influence in CD is stronger compared to UC (Halfvarson 2011, Brant 2011).

3.5.3. Linkage studies in IBD

The familial aggregation and higher prevalence in first-degree relatives of IBD, than in the general population, encouraged the scientific community to look for susceptibility genes enriched in families where multiple members across different generations are affected. As a result of several linkage scans performed in different populations, many genomic regions have been implicated in CD, UC or IBD in general. The results from the first linkage scan (Hugot et al. 1996) were published in 1996 suggesting a putative CD-susceptibility locus on chromosome 16q12 (IBD1). Some studies have confirmed linkage to IBD at this locus (Cho et al. 1997, Curran et al. 1998, Mirza et al. 1998) whereas other studies did not find any evidence of linkage (Olavesen et al. 2000, Rioux et al. 1998). In the same year, the second linkage scan (Satsangi et al. 1996) was performed on 186 affected sibling pairs, belonging to 160 nuclear families. All the study subjects in this study were UK resident Caucasians. The results of this study suggested the presence of additional susceptibility loci for both CD and UC on chromosomes 3p26 (IBD9) and 12p13.2-q24.1 (IBD2). The linkage at IBD9 locus (Curran et al. 1998, Rioux et al. 2000) and IBD2 locus (Hampe et al. 2001, Paavola et al. 2001) was later replicated in independent studies conducted in the European and North American populations. Nevertheless, studies conducted in Canadian (Rioux et al. 1998), Belgian (Vermeire et al. 2000) and French (Lesage et al. 2000) populations reported no evidence for linkage to IBD2 and IBD9 genomic regions raising the possibility that these loci may not be susceptibility loci in all populations. In a study by Cho et al. (174 families and 297 CD, UC or mixed relative pairs), evidence for linkage at chromosomal regions 4q (IBD4), 3q and 1p36 (IBD7) was observed (Cho et al. 1998).
Other regions identified by different linkage scans are the IBD3 locus on chromosome 6p (Hampe et al. 1999b, Hampe et al. 1999a), the IBD5 locus on chromosome 5q31 (Ma et al. 1999)) and the IBD6 locus on chromosome 19p13 (Rioux et al. 2000). The large linkage region on chromosome 16 motivated Hampe et al. to hypothesize that an additional risk gene may exist on this chromosome (Hampe et al. 2002). In a high-density experiment, using 39 microsatellite markers, a distinct locus on chromosome 16p (IBD8) was identified. The genomic regions with linkage to IBD identified in these studies have been replicated in several other studies (Duerr et al. 1998, Ma et al. 1999, Vermeire et al. 2004, Barmada et al. 2004, Paavola-Sakki et al. 2003, Williams et al. 2002, Duerr et al. 2000). However, efforts to find susceptibility genes in the many genomic regions identified by linkage studies have mostly met with disappointment. The linkage studies described here are summarized in Table 4.

Table 4. Linkage studies in IBD.

<table>
<thead>
<tr>
<th>Locus/Chromosome</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBD3 (6p)</td>
<td>(Hampe et al. 1999b, Hampe et al. 1999a)</td>
</tr>
<tr>
<td>IBD4 (4q11-q12)</td>
<td>(Cho et al. 1998)</td>
</tr>
<tr>
<td>IBD5 (5q31)</td>
<td>(Rioux et al. 2000, Ma et al. 1999)</td>
</tr>
<tr>
<td>IBD6 (19p13)</td>
<td>(Rioux et al. 2000)</td>
</tr>
<tr>
<td>IBD7 (1p36)</td>
<td>(Cho et al. 1998)</td>
</tr>
<tr>
<td>IBD8 (16p)</td>
<td>(Curran et al. 1998, Hampe et al. 2002)</td>
</tr>
<tr>
<td>IBD9 (3p26)</td>
<td>(Curran et al. 1998, Rioux et al. 2000)</td>
</tr>
</tbody>
</table>

*underlined: no evidence for linkage*
In contrast to the linkage studies, association studies have been more promising. This approach has led to the identification of several susceptibility loci in IBD. The IBD1 locus on chromosome 16 was identified in the first genome-wide linkage scan (Hugot et al. 1996). In a separate study using a positional-cloning strategy, Hugot et al. identified three independent associations for CD (Hugot et al. 2001). The three associated \textit{NOD2} variants are R702W, G908R and 1007\textit{fs}, where the first two are missense variants and the latter is a frame-shift mutation and the most significant of all three. Independently, Ogura \textit{et al.} reported a cytosine insertion at nucleotide 3020 \textit{(3020insC)} in exon 11 of the \textit{NOD2} gene resulting in a frame-shift at the second nucleotide of codon 1007, as previously reported by Hugot \textit{et al.} (Ogura et al. 2001). The \textit{caspase recruitment domain-containing protein 15/nucleotide-binding oligomerization domain protein 2 (CARD15/NOD2)} gene lies in this genomic region, with mutations in this gene being associated with susceptibility of CD (Ogura et al. 2001). The first GWAS was performed in 2005 in IBD patients from Japan and the United Kingdom (Yamazaki et al. 2005). This study identified variations in the \textit{tumor necrosis factor superfamily, member 15 (TNFSF15)} gene which conferred IBD susceptibility in these populations. This association was later replicated in some studies but could not be established in others (Picornell et al. 2007, Thiebaut et al. 2009). Nonetheless, GWAS proved to be a fairly successful method as several associations were reported in the following years. A highly significant association between the \textit{IL-23 receptor (IL-23R)} gene and CD was reported in a study consisting of cases and controls of European descent (Duerr et al. 2006) and further studies in IBD families revealed that this gene was also associated with UC in non-Jewish Caucasians. This finding has been successfully replicated in different populations (Libioulle et al. 2007, Raelson et al. 2007, Wellcome Trust Case Control Consortium 2007). Other genes which have been reported to be associated with IBD in different CD GWAS are \textit{prostaglandin receptor 4, EP4 subtype (PTGER4)} (Libioulle et al. 2007, Wellcome Trust Case Control Consortium 2007, Rioux et al. 2007), \textit{immunity-related GTPase family M (IRGM)} (Wellcome Trust Case Control Consortium 2007) and \textit{protein-tyrosine phosphatase, non-receptor type 2 (PTPN2)} (Wellcome Trust Case Control Consortium 2007). Together with \textit{IL23R} gene, the \textit{autophagy 16-like 1 (ATG16L1)} gene is the most replicated gene, although the association of \textit{ATG16L1} is mainly with CD and not with UC (Wellcome Trust Case Control Consortium 2007, Hampe et al. 2007). Previous studies in IBD families and twin studies had suggested that the genetic component is stronger in CD. Perhaps for that reason, more GWAS have been performed in CD than in UC. Nevertheless, in contrast to the strong association of \textit{ATG16L1} with CD, a SNP (rs3024505) in the vicinity of IL-10 was strongly associated with UC in a GWAS on multiple populations of European descent (Franke et al. 2008). In another GWAS (Silverberg et al. 2009), loci on chromosome 1p36 and 12q15 showing significant association
with UC were identified. Multiple genes involved in inflammation and immunity are present in this region. In addition, an association between UC and IL23R and the MHC region was independently verified in this study. It was earlier thought that UC might be a milder form of CD and when only some of the CD-specific genes are present UC may result. It has now been recognized that UC and CD are two closely related but distinct disorders of the gut. The two diseases share many risk genes, but also have a number of disease-specific risk genes. In a GWAS comprising 2693 UC patients and 6791 control subjects, all of European ancestry, McGovern et al. identified several novel UC susceptibility loci and confirmed association with previously reported loci (McGovern et al. 2010a). Furthermore, several CD risk genes were also found to be associated with UC. Genome-wide association studies have also been performed to identify genes important in pediatric UC and early onset UC (Imielinski et al. 2009, Kugathasan et al. 2008).

As is the case with complex diseases, many genes contribute to a disease risk but the contribution of a single gene may be small. Meta-analyses have been quite useful in identifying low-risk genes by using the cumulative power of several different GWAS. In a CD meta-analysis, 71 confirmed SNP associations were reported (Franke et al. 2010). A meta-analysis in UC led to an increased number of confirmed SNP associations, from 29 to 47 (Anderson et al. 2011) with a significant overlap with CD. Recently, an imputation-based association analysis of a large number of UC and CD GWAS using 1.23 million SNPs and a subsequent meta-analysis with Immunochip validation was performed (Jostins et al. 2012). This study identified 71 new IBD associations, taking the total number to 163. One hundred ten of the total 163 loci are associated with both CD and UC, 30 loci are CD specific and 23 loci are UC specific. Collectively, these loci explain only a minority of the variance in the disease risk (< 15% for CD and < 8% in UC) (Jostins et al. 2012) suggesting a bigger role for rare genetic variants and other experimental risk factors. A systematic review and meta-analysis of data from 93 previously published studies (in different Asian populations) revealed that there are major differences in IBD risk genes in Western and Asian countries (Ng et al. 2012). For instance, major \textit{NOD}-2 variants that are associated with CD in Western countries are not associated with CD in Asian populations studied in this study. However, other population-specific \textit{NOD}-2 variants were reported to exist. This demonstrates that population-specific differences in risk genes are likely to exist. The association studies described above are tabulated in Table 5.
Table 5. Association studies in IBD.

<table>
<thead>
<tr>
<th>Loci/Genes identified or investigated</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBD1/NOD2 (16q12)</td>
<td>(Hugot et al. 2001, Ogura et al. 2001)</td>
</tr>
<tr>
<td>IL23R (1p31.3)</td>
<td>(Duerr et al. 2006, Libioulle et al. 2007, Raelson et al. 2007, Wellcome Trust Case Control Consortium 2007)</td>
</tr>
<tr>
<td>PTGER4 (5p13.1)</td>
<td>(Libioulle et al. 2007, Wellcome Trust Case Control Consortium 2007, Rioux et al. 2007)</td>
</tr>
<tr>
<td>IRGM (5q33.1)</td>
<td>(Wellcome Trust Case Control Consortium 2007)</td>
</tr>
<tr>
<td>PTPN2 (18p11.3-11.2)</td>
<td>(Wellcome Trust Case Control Consortium 2007)</td>
</tr>
<tr>
<td>ATG16L1 (2q37.1)</td>
<td>(Wellcome Trust Case Control Consortium 2007, Hampe et al. 2007)</td>
</tr>
<tr>
<td>IL10 (1q31-q32), ARPC2 (2q35)</td>
<td>(Franke et al. 2008)</td>
</tr>
<tr>
<td>IBD7 (1p36) and 12q15, multiple genes</td>
<td>(Silverberg et al. 2009)</td>
</tr>
<tr>
<td>Multiple UC susceptibility loci</td>
<td>(Jostins et al. 2012, Anderson et al. 2011, McGovern et al. 2010a)</td>
</tr>
<tr>
<td>Multiple CD susceptibility loci</td>
<td>(Jostins et al. 2012, Franke et al. 2010)</td>
</tr>
</tbody>
</table>

3.6. Pathogenesis of IBD

In the past few years, immense knowledge has been gained regarding genetic and environmental factors important in the onset of IBD, but despite this progress the pathogenesis of IBD remains elusive. There is however a general consensus now that IBD may be a result of an inappropriate and continuous inflammatory response to commensal gut microbes in genetically susceptible individuals. Genes identified in different genetic studies (Jostins et al. 2012, Anderson et al. 2011, Franke et al. 2010, Franke et al. 2008) suggest several cellular pathways that are crucial in intestinal homeostasis including innate and adaptive immune regulation, epithelial restitution, microbial defense, reactive oxygen species (ROS) generation, endoplasmic reticulum (ER) stress and autophagy (Khor, Gardet & Xavier 2011, Maloy, Powrie 2011). The CD susceptibility gene CARD15/NOD2 encodes a protein which belongs to a family of intracellular proteins that contain an N-terminal caspase recruitment domain (CARD), a centrally located nucleotide-binding domain (NBD) and a C-terminal regulatory
leucine-rich repeat (LRR) domain (Ogura et al. 2001). NOD2 acts as an intracellular receptor for bacterial products in different cell types and transduces signals leading to NFkB activation and plays a key role in innate immune activation (Strober et al. 2006, Creagh, O'Neill 2006). The frame-shift mutation-1007fs- results in the Leu1007Pro substitution in the tenth LRR of the NOD2 protein followed by a premature stop codon and a truncated protein (Ogura et al. 2001). Cooney et al. showed that the stimulation of NOD2 by muramyldipeptide induces autophagy in dendritic cells (Cooney et al. 2010). This effect is dependent on, among others, the ATG16L1 protein, which is a product of the CD associated ATG16L1 gene. Furthermore, DCs from CD individuals with NOD2 or ATG16L1 risk variants are defective in autophagy induction, bacterial trafficking and antigen presentation. This further underscores the importance of a dysregulated immune response towards commensal gut flora in IBD pathogenesis (Ordas et al. 2012, Frank et al. 2011).

The inflammation in IBD starts as a result of defects in the components of the innate immune system, but the adaptive immune system is mainly responsible for maintaining the subsequent chronic inflammatory state. In both UC and CD, the balance between Treg cells and effector T cells, i.e. Th1, Th2 and Th17 CD4+ T cells is perturbed (Ordas et al. 2012, Baumgart, Sandborn 2012). In UC, an atypical Th2 type adaptive immune response predominates. This response is mediated by IL-5 and IL-13 secreting non-classical NK cells. Interleukin-13 has cytotoxic effects on IECs, such as the induction of apoptosis and alteration in the TJ proteins (Heller et al. 2005, Heller et al. 2008). In contrast, the mucosal inflammation in CD is a result of Th1 and Th17 immune responses (Hisamatsu et al. 2012, Sakuraba et al. 2009). Several genes that have been identified participate in the regulation of adaptive immune response (Jostins et al. 2012). The delineation of the exact mechanisms leading to impaired gut mucosal homeostasis warrants further research.

3.7. Diagnosis and treatment

According to the guidelines for the diagnosis and management of IBD, issued by the World Gastroenterology Organization, diagnosis of IBD is made on the basis of medical history, physical examination, laboratory and microbial (stool cultures) studies, histological examinations, endoscopy, imaging studies and consultation with other specialists for extraintestinal symptoms. The diagnosis should not be made or excluded on the basis of one variable or result. Other diseases that mimic CD or UC require extra care and must be ruled out. Ulcerative colitis and CD are closely related and may be difficult to differentiate during the initial disease course. However, several features (clinical, endoscopic and radiological, histopathological and serological markers) have been outlined to ease their identification (Bernstein et al. 2010). Once a diagnosis has been established, patients are phenotyped according to the Montreal classification criteria (Silverberg et al. 2005) and screened for
extraintestinal manifestations and associated autoimmune diseases. The treatment regimen in IBD depends on disease course, extent of inflammation and disease severity. However, the general treatment goals are to improve and maintain the general well-being of patients, treatment of acute disease, maintain steroid-free remissions, prevent complications, hospitalizations and surgery and maintain good nutritional status (Bernstein et al. 2010). Drugs that are used in IBD management include aminosalicylates (anti-inflammatory drugs), corticosteroids, immune modifiers, anti-TNF agents and antibiotics. Probiotics, different experimental agents in UC and CD (antiadhesion molecules, anticytokine therapies etc.) and symptomatic therapies may also be used to treat the subject. In about 70%-75% of the CD cases, surgery is required at some point to relieve the symptoms if the drug treatment is unsuccessful. In 25%-30% of UC patients, surgery is performed if the patient is not responding to medical treatment or in the presence of dysplasia. However, in contrast to surgery in CD, surgical resection in UC is considered curative for the disease (Bernstein et al. 2010).

4. Gut microbiota, immune system and susceptibility to different diseases

Our immune system evolved to protect us from harmful pathogens that surround us. Yet, our body is a host to numerous microbes that inhabit the gut and skin surface. It is estimated that the human gut is home to approximately $10^{14}$ microorganisms, which is ten times the number of cells in our body (Bjorksten et al. 2001, Sears 2005). Over the course of human evolution, a symbiotic relationship has evolved between humans and the gut resident microorganisms. The human host provides nutrition and habitat to these microbes and in return they perform various tasks in the gut, which our body is incapable of doing. It is a formidable task to identify all microbial species in the gut owing to inter-individual differences, challenges in their identification or because they cannot be cultured in the lab (Sears 2005, Guarner, Malagelada 2003). In general, human beings have a lot more anaerobic bacteria (100-1000 times) compared to aerobic bacteria (Guarner, Malagelada 2003) and every individual has several hundreds of species belonging to each genera. However, most gut bacteria belong to the genera Bacteroides, Clostridium, Bifidobacterium, Peptostreptococcus, Pepticoccus, Eubacterium, Ruminococcus and Fusobacterium (Guarner, Malagelada 2003, Beaugerie, Petit 2004). The subdominant genera include aerobic (or facultative anaerobic) bacteria such as escherichia, enterobacter, enterococcus, klebsiella, lactobacillus and proteus among others. Tap et al. studied the composition of gut microbiota in fecal samples from 17 healthy individuals (Tap et al. 2009). They found that a small number of operational taxonomic units (OTUs) were present in over fifty percent of studied individuals and suggested that a core gut microbiota may exist in all individuals. The microbiota composition of newborns differ from that of grown-ups. At birth, the
infant gut is mainly colonized with facultative aerobic bacteria, including streptococci and *Escherichia coli*, which at weaning shift towards obligate anaerobic composition of genera *Bifidobacterium*, *Bacteroides*, *Clostridium* and *Ruminococcus* (Sears 2005).

Microorganisms perform an array of functions for their host. For example bacteria ferment the non-digestible complex dietary carbohydrates and mucus produced by epithelial cells into short chain fatty acids (SCFAs) (Sears 2005, Guarner, Malagelada 2003). This digestion process is aided by different enzymes collectively present in distinct microbial species in the colon. The SCFAs are a major source of useful energy and nutrition for the host and also help in the absorption of dietary minerals like calcium, magnesium and iron (Guarner, Malagelada 2003). In addition to its role in nutrition management, the microbiota is important for the maintenance of epithelial homeostasis. Germ-free rats have fewer cells in the crypts compared to rats housed under normal conditions, a feature which has been attributed to the lack of normal commensal bacteria (Gordon et al. 1997, Alam, Midtvedt & Uribe 1994). The SCFAs produced by bacteria during carbohydrate metabolism stimulate epithelial cell proliferation and differentiation in the small- and large intestine and aid in intestinal homeostasis (Frankel et al. 1994, Siavoshian et al. 2000).

Microorganisms are present in great numbers in the gut so they must compete with each other for an attachment site or nutrients. Such a competition is known as a barrier effect which serves to protect the host from colonization by pathogenic or opportunistic bacteria (Guarner, Malagelada 2003). There is some evidence that commensal bacteria can inhibit the growth of their competitors by producing toxins, called bacteriocins (Lievin et al. 2000). It is thus quite evident that a harmonious relationship with gut microbiota is crucial to our well-being. A disturbance in this relationship usually leads to health complications for the host. In case of a bacterial translocation, bacterial toxins or viable bacteria can cross the intestinal epithelial barrier and spread throughout the body and cause sepsis, shock, multisystem organ failure or the death of the host (Guarner, Malagelada 2003). Bacterial translocation is caused by three main mechanisms, namely overgrowth of bacteria in the small intestine, increased permeability of the intestinal epithelial barrier and deficiencies or defects in the host immune system. In human disorders such as multisystem organ failure, acute severe pancreatitis, advanced liver cirrhosis, intestinal obstruction and IBD, rates of positive mesenteric lymph node culture are 16%-40%, which in healthy individuals are below 5% (Lievin et al. 2000).

Studies in animal models show that gut microbiota is essential for proper development of the host immune system. Mice kept under germ-free conditions have an underdeveloped mucosal immune system with small lymphoid follicles, few IgA-secreting plasma cells and reduced submucosal T-cell populations, including CD4+ and intraepithelial CD8+ T lymphocytes (Macpherson, Uhr 2004a,
Furthermore, studies on germ-free mice and specific-pathogen-free (SPF) mice demonstrate impaired ability of germ-free mice to clear infection compared to SPF mice (Inagaki et al. 1996). Rakoff-Nahoum et al. showed that toll-like receptors (TLRs), which play an important role in protection from infections caused by microbes, recognize and interact with commensal bacteria (Rakoff-Nahoum et al. 2004). This interaction is critical to host protection from gut injury and associated mortality. Moreover, in a study by Ivanov et al. it was shown that specific intestinal microbiota directs the differentiation of IL-17 producing T-helper cells in small-intestinal mucosa and thus regulates the balance between Th17 and Treg cells in the lamina propria (Ivanov et al. 2008). Consequently, gut microbiota influences intestinal immunity and susceptibility to certain diseases. The modulation of intestinal microbiota was shown to delay the onset of T1D, and also causing a reduction in the incidence of T1D in germ-free non-obese diabetic mice (King, Sarvetnick 2011).

The interaction between the host immune system and commensal bacteria is particularly important in childhood or infancy when the immune system is developing. Interestingly, the gut microbiota plays an important role in modulating the adaptive immune system (Macdonald, Monteleone 2005). The GI tract plays an important role in the extrathymic development of T and B cell repertoires. Developing T cells and B cells are exposed to self-antigens in the thymus and bone-marrow respectively which ensures the induction of tolerance towards self-antigens. Similarly, the GI tract is the largest source of non-self, but innocuous antigens, derived from commensal bacteria. Dendritic cells have been shown to sample luminal bacterial antigens and present them to submucosal macrophages and T lymphocytes (Rescigno et al. 2001, Macpherson, Uhr 2004b). This process is considered important in the induction of tolerance towards harmless environmental antigens. It is thus not surprising that an alteration in the gut microbiota has been reported in immunological diseases such as IBD (Andoh et al. 2011, Thomazini et al. 2011, Sokol et al. 2006, Macia et al. 2012), CelD (Macia et al. 2012, Schippa et al. 2010, Nistal et al. 2011), T1D (King, Sarvetnick 2011, Macia et al. 2012, Hara et al. 2013), hepatitis B virus-induced chronic liver disease (Xu et al. 2011), RA and asthma (Macia et al. 2012). An allergic reaction is a result of an overreaction of the immune system to innocuous environmental antigens. Reduced diversity in the composition of gut microbiota during infancy and childhood was found to be associated with increased risk for allergies at later ages (Bjorksten et al. 2001, Johansson et al. 2011, Bisgaard et al. 2011). In a study, an early colonization with a group of lactobacilli was able to decrease the risk for allergy in five-year old children, despite genetic predisposition for allergy in these children (Johansson et al. 2011). In another study, early intervention with probiotics and prebiotics did not affect the risk for airway inflammation later in childhood (Kukkonen et al. 2011). Alterations in the gut microbiota have been reported in systemic
diseases such as obesity (Million et al. 2011) and type-2 diabetes (T2D) (Brown et al. 2012) which are primarily metabolic disorders. This suggests that changes in gut microbiota may have implications beyond immunological diseases.

It was earlier thought that the composition of intestinal microbiota stayed relatively stable after colonization in early childhood. However, several studies now speak against this concept and a clear picture of dietary factors that can affect intestinal microbial composition is starting to emerge. The colonization of the GI tract with microbes in newborns starts already at birth and its composition is highly dependent on the mode of delivery. There are significant differences in the microbiota composition in infants who were delivered through birth canal compared to those who were born through caesarean section (Guarner, Malagelada 2003, Penders et al. 2006). Other factors that affect the diversity of gut microbiota are the type of diet (i.e. breast feeding or formula feeding), age, infant hospitalization, stress, degree of hygiene and antibiotic use by the infant (Brown et al. 2012, Penders et al. 2006). Furthermore, diet rich in fat and sugar results in microbiota which is distinct from microbiota promoted by a diet which is rich in carbohydrate and low in fat and sugar (Brown et al. 2012).

As assessed by a study in mice, the gut microbiota is to a large extent (57%) determined by diet and, interestingly, also to a smaller extent (12%) by host genetics (Zhang et al. 2010). A potential relationship between host genetics and gut microbiota was suggested already in the early 1980s (Wensinck et al. 1981, Van de Merwe, Stegeman & Hazenberg 1983). These studies reported that the fecal microbiota of CD patients was different from that of healthy subjects in composition and that there was a significantly higher degree of similarity in fecal microbiota of healthy monozygotic twin pairs compared to that of dizygotic twin pairs. A more recent study demonstrated a higher degree of fecal microbiota similarity between healthy monozygotic twins than in monozygotic twins with CD (Dicksved et al. 2008). A significant difference in microbiota composition between healthy individuals and patients with ileal CD has also been reported. In another study, Stewart et al. investigated the genetic basis for predominance of eubacteria in fecal microbiota in twin pairs (Stewart, Chadwick & Murray 2005). The highest degree of similarity was found between monozygotic twins followed by dizygotic twins and the least amount of similarity was found between unrelated control individuals. Although many family and twin studies established the involvement of genetic component in microbiota composition, fewer studies have identified genes involved in microbial colonization.

Mutations in the NOD2/CARD15 gene are associated with increased risk for CD and have also been associated with a shift in microbial composition in the ileum (Li et al. 2012). Furthermore, genetic
variants in the genes TLR4 and CD14 have been shown to confer increased risk for CD (Gewirtz et al. 2006) and atopic diseases (Penders et al. 2010). The HLA region is of utmost importance in CelD and HLA-DQ2 genotype has been reported to influence the composition of commensal bacteria, both in humans and mice (Sanchez et al. 2011, Toivanen, Vahtovuo & Eerola 2001). The \textit{fucosyltransferase 2 (FUT2)} gene is present on 19q13.33 and is associated with both alterations in the microbiota and risk or protection from various diseases. The product of the \textit{FUT2} gene is an enzyme (α-1,2-fucosyltransferase) which participates in the conversion of type 1 precursor to type 1 H antigen by adding a terminal α1,2-linked fucose to galactose on type 1 precursor. The type 1 H antigen is the precursor for the further synthesis of A and B blood group antigens which is carried out by A- and B-transferases respectively (Combs 2009).

\begin{center}
\includegraphics[width=\textwidth]{figure5.png}
\end{center}

\textit{Figure 5.} The figure shows the biosynthesis of A, B and H blood group antigens.

These antigens are also expressed in the GI mucosa where they serve as anchors for various microbes. In 20% of Caucasians, this gene is rendered nonfunctional by a nonsense mutation at nucleotide 428, introducing a premature stop codon at position 143 (G428A, W143X) (Kelly et al. 1995, Soejima, Koda 2005). Individuals who are homozygous for this mutation are called non-secretors (se428, rs601338-AA). Although the prevalence of non-secretor status is fairly similar among different populations, different mutations leading to the non-secretor status have been identified (Soejima, Koda 2005). For example, the allele se428 (G428A, W143X) is the causative variant in Europeans, Iranians and Africans, whereas another allelic variant se385 (A385T,
rs1047781) is associated with the non-secretor status in South-East and East-Asians (Ferrer-Admetlla et al. 2009). As FUT2 gene controls the expression of ABH and Lewis histo-blood group antigens, it is conceivable that it may also affect the composition of microbiota. In a study by Wacklin et al., a significant reduction in bifidobacterial diversity, richness and abundance was observed in fecal microbiota from healthy non-secretors compared to secretors (Wacklin et al. 2011). Consistent with the role of gut microbiota in modulating innate- and adaptive immune responses, both secretor and non-secretor status are associated with increased risk for or protection from certain diseases. For example, non-secretor individuals show increased susceptibility to duodenal ulcers (Evans et al. 1968), rheumatic fever (Haverkorn, Goslings 1969) and cholera (Chaudhuri, DasAdhikary 1978), but they are resistant to infections from Norovirus (Lindesmith et al. 2003, Carlsson et al. 2009) and respiratory viruses (Raza et al. 1991). Similarly, secretors are more susceptible to Helicobacter pylori infection (Linden et al. 2008) but the breast milk of secretor mothers confers protection against Campylobacter jejuni to the offspring (Ruiz-Palacios et al. 2003). Furthermore, FUT2 non-secretor status has been associated with increased risk for CD (McGovern et al. 2010b) and T1D (Smyth et al. 2011) in Caucasians. In a study by Rausch et al., which included healthy control individuals and individuals affected with CD, the FUT2 genotype was found to affect the microbiota associated with colonic mucosa (Rausch et al. 2011). In light of these studies, it seems tempting to assume that both environmental factors, such as diet and life style, and the genetic architecture of an individual define the composition of gut microbiota.

5. Transcriptomics: scope and methods

The information required for proper functioning of an organism is encoded in its genome. This information is extracted by gene transcription and expressed through the process of gene translation. The transcriptome is the complete set of RNA [messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA) and other non-coding RNA] transcripts produced by gene transcription at any given time. During an organism’s life time, the genomic content remains largely unchanged and does not vary according to the cell type, stage of development or environmental conditions. In contrast, the transcriptome is highly dynamic and its composition may differ in different cells or in the same cell at different time points or under different conditions. The study of transcriptome produced by the genome at a given time is referred to as transcriptomics. Transcriptomics can be used to investigate differential gene expression patterns in cells at different developmental stages (Assou et al. 2011), to unravel the gene interaction network (Alekseev et al. 2009), to discover genes abnormally expressed in a disease (Wang et al. 2013, Min et al. 2009), to identify drug targets (Cheung 2007) etc. Several hybridization-based methods (Northern blotting,
DNA microarrays), polymerase chain reaction (PCR)-based methods (differential display and representational difference analysis) and sequence-based techniques (expressed sequence tags (ESTs), serial analysis of gene expression (SAGE)) have been used in gene expression studies (Kozian, Kirschbaum 1999, Fryer et al. 2002). Although all these methods have inherent advantages and limitations (relating to expense, technical difficulty, sensitivity and specificity), the microarray technology has been widely used. Recent technological advances have also made possible the use of sequencing techniques, such as exome sequencing (Ng et al. 2009) and RNA sequencing (Wang, Gerstein & Snyder 2009) in gene expression studies. Nevertheless, these methods are still quite expensive and due to a large amount of data produced also present with some technical and bioinformatics-related challenges.

6. Several risk genes are shared among different immunological diseases

As discussed in sections 2.6 and 2.8, individuals suffering from CelD and DH have an increased risk for developing other autoimmune diseases. This suggests that same genes or same pathways are responsible for causing different autoimmune/immune-mediated diseases. Indeed, recent research has shown that many genes are shared across different immunological diseases. Susceptibility genes in CelD have been shown to also confer risk to other autoimmune diseases such as T1D (Smyth et al. 2008), RA (Coenen et al. 2009) and CD (Festen et al. 2011). Among many others, shared genetic background between RA and T1D (Alizadeh et al. 2007, Zhernakova et al. 2007, Eyre et al. 2010), between CD and psoriasis (Ellinghaus et al. 2012) and between CD and UC (Jostins et al. 2012, Glas et al. 2012, Waterman et al. 2011) has been reported. Cotsapas et al. studied the extent of sharing for 107 immune-disease risk SNPs in seven immunological diseases (Cotsapas et al. 2011). They found that 47 of 107 disease risk loci were associated with multiple but not all diseases that were under investigation. A detailed overview of shared genetics between IBD and several other diseases has been provided in a review by Lees et al. (Lees et al. 2011). Undoubtedly, to look for shared genetics between similar diseases is an effective method as it is both economical and time saving.
Aims of this study

CelD and IBD are inflammatory disorders where both genetic and environmental factors contribute to disease pathogenesis. This study was conducted to identify genes which are affected by environmental factors (gliadin peptides) as well as to identify genes which affect environmental factors (gut microbiota). Furthermore, genes involved in the regulation of innate and adaptive immune system are likely to be shared by different immunological diseases. Both CelD and IBD are immunological diseases of the gut, efforts were made to identify CelD genes which may also confer susceptibility to IBD in the Finnish population. The specific aims of this study were:

1. To use microarray chips to obtain a gene expression profile of intestinal epithelial Caco-2 cells stimulated with gliadin in order to gain insight into the activation of different receptor-dependent or –independent pathways. These pathways may help us understand the gliadin mediated early steps in disease progression.

2. To study the association of secretor/non-secretor status (FUT2 rs601338 polymorphism) with CelD and IBD in the Finnish population. The FUT2 genotype affects the composition of gut microbiota and may therefore associate with the risk of these diseases.

3. To explore the shared genetic background of CelD and IBD by investigating whether the recently identified CelD risk loci also affect susceptibility of IBD.
Material and methods

1. Digestion of gliadin, bovine serum albumin and lactalbumin

Wheat flour (Raisio Oyj, Raisio, Finland) was obtained from a local grocery store. Bovine serum albumin (BSA), lactalbumin, pepsin (P-6887) and trypsin (T-7418) were obtained from Sigma-Aldrich (Sigma-Aldrich, Seelze, Germany). The method for the extraction of gliadin from wheat and for the digestion of gliadin, BSA and lactalbumin with pepsin and trypsin is described below. At the same time and using the same reagents, a blank pepsin and trypsin control (PT) was prepared where no substrate was used for digestion. For simplicity, the method below describes the digestion of gliadin only. The pepsin- and trypsin-digested gliadin, BSA and lactalbumin are denoted PT-G, PT-BSA and PT-L.

Whole gliadin was first extracted from wheat flour. To remove salt-soluble proteins, 10g of wheat flour was dissolved in 30ml 1M NaCl and incubated at room temperature for one hour on a shaker. This was followed by centrifugation at 4000g for 20 minutes. The pellet was washed with 40ml water and centrifuged at 4000g for 20 minutes. The pellet was then suspended in 25ml 70% ethanol followed by incubation at 60°C for one hour with shaking. After one hour, supernatant containing the gliadin fraction was collected by centrifugation, frozen and lyophilized.

For digestion by pepsin, 60mg gliadin was dissolved in 10ml 50mM NaAc buffer (pH 4.0) and 3mg pepsin (P-6887, Sigma-Aldrich) was added to the solution. The mixture was incubated at 37°C for two hours on a shaker. Seventy one milligrams of Na$_2$HPO$_4$ was added to the solution and the pH was adjusted to 7.0 with NaOH. For digestion with trypsin, 3mg trypsin (T-7418, Sigma-Aldrich) was added to the mixture followed by incubation at 37°C for two hours on a shaker. The digestion solution was heated at >95°C for ten minutes to stop digestion reactions. The pepsin- and trypsin-digested gliadin (PT-G) mix was then frozen, lyophilized and stored at -20°C for later use.

2. Caco-2 cell cultures and stimulations

The human colon carcinoma cell line, Caco-2, was used to study the effect of gliadin peptides on small-intestinal epithelial cells. Despite the colonic origin of Caco-2 cells, when cultured under specific conditions their phenotype, morphologically and functionally, resembles that of enterocytes lining the small intestine (Hidalgo, Raub & Borchardt 1989). Caco-2 cells were maintained in minimum essential medium (MEM, Invitrogen, Carlsbad, CA, USA) to which 1X sodium pyruvate (NaPy 100X, Invitrogen), 1% penicillin-streptomycin (Pen-Strep, Invitrogen), 1X non-essential amino acids (100X NEAA, Invitrogen) and 10% fetal calf serum (FCS, PROMOCCELL GmbH, Heidelberg,
Germany) were added. The cells were kept in an incubator at 37°C with 5% carbon dioxide. The pepsin- and trypsin (PT) blank control, PT-BSA, PT-L, PT-G and the synthetic peptides p31-43 (LGQQQPFPPQYPY) and p57-68 (QLQPFPQPQLPY) (both obtained from New England Peptide LLC, Gardner, MA, USA) were dissolved in MEM + 1% FCS and sterilized by passing through 0.22µm pore filter (Millipore, Billerica, MA, USA). An endotoxin test (Limulus Amebocyte Lysate (LAL), Pyrogent plus single test kit, CAMBREX, East Rutherford, NJ, USA) confirmed the presence of endotixin in PT-BSA, but not in PT-G. For stimulation, 3x10^5 cells were seeded in 3ml MEM + 1% FCS in 25cm^2 flasks at 37°C in an incubator with 5% carbon dioxide for three days. Before stimulations, the cells were washed once with phosphate-buffered saline (PBS) and stimulated with PT, PT-G, PT-BSA, PT-L, p31-43 and p57-68 peptides in 2ml MEM + 1% FCS for six hours. The final concentration of PT, PT-BSA, PT-L and PT-G to which the cells were exposed was 1mg/ml, whereas that of synthetic peptides (p31-43 and p57-68) was 150µg/ml. PT, PT-BSA and PT-L were used as negative controls for PT-G treatment and p57-68 peptide was used as a negative control for p31-43 stimulation. The control cells (MED-CTL) were kept in MEM + 1% FCS for six hours. The experiment was repeated four times to obtain four different biological replicates. The cells differed by 1-2 passages between each experiment.

3. Cell collection and RNA extraction

After the six hours stimulation time, cells were collected by adding 1ml 1X trypsin-EDTA (Trypsin 10X, Invitrogen) to the flasks followed by incubation at 37°C for four to five minutes. Trypsin was inhibited by adding 1ml MEM + 10% FCS to the flasks followed by centrifugation at 9600xg for five minutes. The supernatant was discarded and the pellet was washed once with 500µl 1X PBS followed by another round of centrifugation. The supernatant was discarded and the pellet was stored at -80°C until later use. For RNA extraction, the pellet was thawed on ice and RNA extracted using RNAeasy Plus Mini kit (Qiagen GmbH, Hilden, Germany) following the vendor’s instructions. RNA quality was assessed by bioanalyzer (2100 Bioanalyzer, Agilent, Santa Clara, CA, USA).

4. Microarray hybridizations and data analysis

RNA extracted from cells treated with PT, PT-G, MED-CTL and p31-43 peptide (from four different experiments) was used for hybridization on Agilent one color human whole genome DNA microarray chips (4x44K) (Agilent Technologies, Santa Clara, CA, USA). 600µg RNA was used for cDNA synthesis. The cDNA synthesis and subsequent hybridization on microarray chips were carried out at Biomedicum Genomics Support, Helsinki, Finland. Microarray raw data files were imported into R v. 2.15 (Team 2012) and analyzed with BioConductor (Gentleman et al. 2004) package limma (Smyth 2005). After the quality check, the microarray probes were filtered and re-annotated as previously
described (Gertz et al. 2009) and their median foreground intensity was normalized with the quantiles method without applying any background correction, as suggested by Zahurak et al. (Zahurak et al. 2007). After that, the probes were tested for differential expression using a linear model followed by a moderated t-test (Smyth 2005) for intended comparisons. All genes with p < 0.01 were considered to be differentially expressed.

5. Quantitative real-time polymerase chain reaction (qRT-PCR)

Five genes differentially regulated in microarray data analysis, were selected for validation by qRT-PCR. The selection criteria were high endogenous expression level (≥ log2 9.0) and a difference in fold change (FC) of ≥ 1.4 in either direction. One probe corresponding to the gene reticuloendotheliosis viral oncogene homolog A (RELA) did not fulfill the selection criteria. This gene was selected because of its involvement in the NFkB pathway and NFkB pathway is relevant in CelD pathogenesis (Maiuri et al. 2003b). Primers were designed using the primer designing software Primer express (Applied Biosystems Foster city, CA, USA) and are listed in Table 6. Complementary DNA (cDNA) was made from 300ng of RNA using reagents provided in the TaqMan Reverse Transcription Kit (Applied Biosystems, Foster city, CA, USA) and following the enclosed instructions. The qRT-PCR assay was performed on 7500 Fast Real-Time PCR System (Applied Biosystems, Foster city, CA, USA) using Power SYBR Green PCR Master Mix (Applied Biosystems) (1X SYBR Green PCR Master Mix and 0.6µM forward and reverse primer each). The qRT-PCR data were analyzed using the comparative C(T) method as described previously (Schmittgen, Livak 2008). The gene expression levels were normalized against beta-2 microglobulin (B2M) gene expression.

Table 6. Primers used for qRT-PCR

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward (5’-3’)</th>
<th>Reverse (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2M</td>
<td>GTGCTCGCGCTACTCTCTC</td>
<td>GTCAACTTCAATGTCGGAT</td>
</tr>
<tr>
<td>CEBPa</td>
<td>GACCCTCACCTTTGTGGTACTG</td>
<td>CTGATCGTGCTTGTTTCTCCT</td>
</tr>
<tr>
<td>KLF4</td>
<td>ACCAGGCACCTACGTAACACACA</td>
<td>GCTCGTGCACATTTTGG</td>
</tr>
<tr>
<td>MNF1</td>
<td>CACAGACACCTGGAGAGCTTAA</td>
<td>GGGCAAACTTCTCTGAGTT</td>
</tr>
<tr>
<td>RELA</td>
<td>CAGGCGAGAGGAGCCAGACA</td>
<td>TGTGTAGCCATTGATTTGATGG</td>
</tr>
<tr>
<td>SPRED1</td>
<td>TGAATGCTGTCATACAACAGATGTC</td>
<td>GAAAGTTACAGTCTTTCTAGGAACCA</td>
</tr>
</tbody>
</table>

6. DNA extraction

Prior to the studies described here the genomic DNA was extracted from the EDTA whole blood samples of the CelD and IBD patients using the Flexigene DNA kit or the QIAamp DNA Blood Mini Kit (Qiagen GmbH, Hilden, Germany).

7. Marker selection

Association of only one SNP (rs601338-FUT2) was studied in Study II. The selection of this SNP was based on its association with CD in another Caucasian population (McGovern et al. 2010b) and also because genetic variants at this position affect the transcription of the FUT2 gene. In Study III, association of 45 SNPs with CD and/or UC was investigated based on their association with CelD in a previous multi-population CelD GWAS (Dubois et al. 2010). Thirty-nine of these SNPs were chosen because they represented the strongest associations from that GWAS. Six additional SNPs were included in Study III as they showed stronger association with CelD in the Finnish population at these loci, than the markers giving the top association signals in the multi-population analysis. All the SNPs investigated during this thesis work are listed in Table 7.

8. Study subjects

The collection of patient and control materials was approved by the ethics committee of Helsinki University Hospitals, Finnish National Public Health Institute and Tampere University Hospital. All the study subjects were informed about the study according to the study protocol and gave their written informed consent. In Study II, there were 943 unrelated CelD patients and an independent set of 119 CelD patients who had DH (Table 8). Most of the CelD and DH patients in this study derive from a large family collection which includes probands and their first-degree relatives screened serologically for CelD associated autoantibodies (ATGA and/or EMA). Ninety-five percent of the Ab-positive CelD family members were also biopsy-proven CelD cases. Only one patient per family, the proband in most cases, was included in this study. Only unrelated individuals were included in the combined CelD + DH dataset. If a CelD patient and a DH patient belonged to the same family, one of them was randomly selected. Hence, the total number of patients in CelD + DH group is 1025.

In the IBD dataset used in this study, there were 303 CD and 545 UC patients (Table 8). Thirty-one of the UC and CD affected individuals were related, thus one of the two was randomly selected to the IBD (CD + UC) group. After randomly removing one of the related UC and CD patients, there were a total of 745 unrelated individuals in the combined dataset group (i.e. IBD group). One thousand forty-five (1045) Finnish blood donor control samples were provided by the Finnish Red Cross Blood Service (Study II) and data for another 1693 control subjects were obtained from the second GWAS.
in CelD by Dubois et al. (Dubois et al. 2010) (Studies II and III). The total number of controls in this study was thus 2738.

Table 7. List of SNPs used in Study II and Study III.

<table>
<thead>
<tr>
<th>MARKER</th>
<th>CHR</th>
<th>POS (Mb)</th>
<th>MARKER</th>
<th>CHR</th>
<th>POS (Mb)</th>
<th>MARKER</th>
<th>CHR</th>
<th>POS (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3748816</td>
<td>1</td>
<td>1.8</td>
<td>rs13098911</td>
<td>3</td>
<td>46.3</td>
<td>rs9792269</td>
<td>8</td>
<td>124.6</td>
</tr>
<tr>
<td>rs12727642</td>
<td>1</td>
<td>7.2</td>
<td>rs6806528</td>
<td>3</td>
<td>69.3</td>
<td>rs1250552</td>
<td>10</td>
<td>75.1</td>
</tr>
<tr>
<td>rs426561*</td>
<td>1</td>
<td>7.2</td>
<td>rs11712165</td>
<td>3</td>
<td>116.5</td>
<td>rs11221332</td>
<td>11</td>
<td>124.3</td>
</tr>
<tr>
<td>rs10903122</td>
<td>1</td>
<td>23.6</td>
<td>rs17810546</td>
<td>3</td>
<td>157.1</td>
<td>rs653178</td>
<td>12</td>
<td>109.0</td>
</tr>
<tr>
<td>rs6691768</td>
<td>1</td>
<td>59.9</td>
<td>rs10936599</td>
<td>3</td>
<td>166.9</td>
<td>rs2762051</td>
<td>13</td>
<td>31.6</td>
</tr>
<tr>
<td>rs864537</td>
<td>1</td>
<td>138.7</td>
<td>rs1464510</td>
<td>3</td>
<td>185.5</td>
<td>rs4899260</td>
<td>14</td>
<td>49.4</td>
</tr>
<tr>
<td>rs859637</td>
<td>1</td>
<td>143.9</td>
<td>rs13151961</td>
<td>4</td>
<td>118.8</td>
<td>rs12928822</td>
<td>16</td>
<td>11.3</td>
</tr>
<tr>
<td>rs2816316</td>
<td>1</td>
<td>163.8</td>
<td>rs13119723*</td>
<td>4</td>
<td>118.9</td>
<td>rs2074404</td>
<td>17</td>
<td>40.2</td>
</tr>
<tr>
<td>rs296547</td>
<td>1</td>
<td>172.1</td>
<td>rs13109373*</td>
<td>4</td>
<td>158.8</td>
<td>rs1893217</td>
<td>18</td>
<td>12.8</td>
</tr>
<tr>
<td>rs13003464</td>
<td>2</td>
<td>60.9</td>
<td>rs1033180</td>
<td>6</td>
<td>0.3</td>
<td>rs601338</td>
<td>19</td>
<td>48.4</td>
</tr>
<tr>
<td>rs17035378</td>
<td>2</td>
<td>68.3</td>
<td>rs10806425</td>
<td>6</td>
<td>88.1</td>
<td>rs242372*</td>
<td>21</td>
<td>14.3</td>
</tr>
<tr>
<td>rs917997</td>
<td>2</td>
<td>96.8</td>
<td>rs802734</td>
<td>6</td>
<td>125.9</td>
<td>rs4819388</td>
<td>21</td>
<td>31.0</td>
</tr>
<tr>
<td>rs13010713</td>
<td>2</td>
<td>173.9</td>
<td>rs2327832*</td>
<td>6</td>
<td>135.5</td>
<td>rs2298428</td>
<td>22</td>
<td>4.9</td>
</tr>
<tr>
<td>rs4675374</td>
<td>2</td>
<td>196.6</td>
<td>rs9295089*</td>
<td>6</td>
<td>156.9</td>
<td>rs5979785</td>
<td>X</td>
<td>10.7</td>
</tr>
<tr>
<td>rs4675377*</td>
<td>2</td>
<td>196.7</td>
<td>rs1738074</td>
<td>6</td>
<td>156.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs13314993</td>
<td>3</td>
<td>33.0</td>
<td>rs6974491</td>
<td>7</td>
<td>37.3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The SNP positions are based on Genome build 37.1 HuRef. The SNP in bold text was used in Study II and all other SNPs were used in Study III. *SNP not from the CelD GWAS, x: SNP genotyped by Taqman.

Table 8. The number of subjects (identified by group) in Study II and Study III are tabulated.

<table>
<thead>
<tr>
<th>Group</th>
<th>Study II</th>
<th>Study III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cohort_FIN</td>
<td>Cohort_FIN</td>
</tr>
<tr>
<td>Controls</td>
<td>2738</td>
<td>2482</td>
</tr>
<tr>
<td>CelD</td>
<td>943</td>
<td>119</td>
</tr>
<tr>
<td>DH</td>
<td>119</td>
<td></td>
</tr>
<tr>
<td>CelD + DH</td>
<td>1025</td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>303</td>
<td>240</td>
</tr>
<tr>
<td>UC</td>
<td>545</td>
<td>459</td>
</tr>
<tr>
<td>IBD</td>
<td>745</td>
<td>699</td>
</tr>
</tbody>
</table>

Study II: rs601338-FUT2 association study; Study III: Shared CelD-IBD genes; Cohort_FIN: Finnish study subjects; Cohort_SWE: Swedish study subjects
Patient and control samples from Finland and Sweden were included in Study III (Table 8). In the Finnish dataset, there were 699 unrelated IBD patients. Of these, 240 patients had CD and 459 patients had UC. There were 2482 healthy controls in study III derived from the FINRISK and Health 2000 cohorts (FINRISK, Health 2000), already genotyped in the CelD GWAS (Dubois et al. 2010) for the studied markers. The IBD patients in both studies (Study II and Study III) were recruited at the Helsinki University Hospital and Maria Hospital and also at University Hospitals in Turku, Tampere, Kuopio and Oulu. The diagnostic criteria and methods of clinical assessment of these patients have been described previously (Helio et al. 2003, Halme et al. 2002). The Swedish sample collection consisted of 923 UC cases and 341 healthy controls and has been described in detail in a previous study (McGovern et al. 2010a). To study the association of CelD risk genes with different CD and UC subphenotypes (Table 9), all the patients were divided into groups based on the clinical presentation of their disease. For every subphenotype that was studied, the patients were regrouped accordingly, thus it was possible for one patient to be present in different groups. In brief, the CD patients were divided into 16 groups and the UC patients were divided into six groups. These groups are summarized in Table 9.

9. Genotyping

The FUT2 SNP marker rs601338 was genotyped in 1045 controls, 943 CelD patients, 119 DH patients, 303 CD patients and 545 UC patients using TaqMan chemistry (Life technologies, Carlsbad, CA, USA) with a forward primer (GGGAGTACGTCCGCTTCAC), a reverse primer (TGCGGGAGGTGGTGGTA), reporter 1-VIC (CTGCTCCTAGACCTT) and reporter 2-FAM (CTGCTCCTGGACCTT). All the patient samples were genotyped in house using ABI 7900HT real-time PCR system. The blood donor control samples (N=1045) had been earlier genotyped at an external facility (Technology Centre, Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland) using the same TaqMan primers and probes and a LightCycler 480 Real-Time PCR Instrument (Roche, Basel, Switzerland). The software version 1.5 (endpoint genotyping analysis) was used to call the genotypes for control samples, and the success rate was 100%. The genotyping was successful for 879 of 943 CelD patients, 116 of 119 DH patients, 280 of 303 CD patients and 496 of 505 UC patients giving a success rate of 93%, 97.5%, 92.4% and 91% respectively. The genotyping success rate for all the cases together was 92.8%. The genome-wide data from previous GWAS (Dubois et al. 2010) (Illumina 610-Quad platform) was used to impute genotypes for 1693 control subjects using the IMPUTE2 program (Howie, Donnelly & Marchini 2009). For genotype imputations, the 1000 genomes phase I interim build (June 2011) was used as reference (1000 Genomes Project Consortium 2010) and only genotypes with a quality score ≥ 0.95 were used in the analyses. To check the reliability of this
method, genotypes for 609 of 943 directly genotyped CeID patients were obtained by imputing. The concordance rate between the two methods was 97.7%. For 30 CeID patient samples, also previously genotyped genome wide, direct genotyping by Taqman chemistry had failed, hence their imputed genotypes were used in the analysis. Thus, genotypes for 909 (879 + 30) CeID patients were available for the analysis. The distribution of genotypes in control datasets did not deviate significantly from Hardy-Weinberg equilibrium (P > 0.05).

Table 9. The subphenotypes of CD and UC patients along with the number of affected individuals in each group are given below.

<table>
<thead>
<tr>
<th>Patient subphenotype</th>
<th>Cohort_FIN</th>
<th>Cohort_SWE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crohn’s disease</td>
<td>n = 240</td>
<td></td>
</tr>
<tr>
<td>Onset age: 0-18 yrs</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Onset age: 19 yrs and above</td>
<td>196</td>
<td></td>
</tr>
<tr>
<td>Familial CD</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Sporadic CD</td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>Left or Right colitis</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Extensive colitis</td>
<td>113</td>
<td></td>
</tr>
<tr>
<td>Terminal ileitis (L1)</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>Affected colon (L2)</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Ileocolonic disease (L3)</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>Perianal fistula</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>No Perianal fistula</td>
<td>193</td>
<td></td>
</tr>
<tr>
<td>Stricturing phenotype</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>Penetrating and Fistulising phenotype</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Non-stricturing Non-penetrating phenotype</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>Mutations in NOD2</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>No mutations in NOD2</td>
<td>196</td>
<td></td>
</tr>
<tr>
<td>Ulcerative colitis</td>
<td>n = 459</td>
<td>n = 923</td>
</tr>
<tr>
<td>Onset age: 0-18 yrs</td>
<td>59</td>
<td>61</td>
</tr>
<tr>
<td>Onset age: 19 yrs and above</td>
<td>382</td>
<td></td>
</tr>
<tr>
<td>Familial UC</td>
<td>115</td>
<td>51</td>
</tr>
<tr>
<td>Sporadic UC</td>
<td>326</td>
<td>483</td>
</tr>
<tr>
<td>Left-sided colitis</td>
<td>166</td>
<td></td>
</tr>
<tr>
<td>Extensive colitis</td>
<td>275</td>
<td></td>
</tr>
</tbody>
</table>

Cohort_FIN: Finnish study subjects; Cohort_SWE: Swedish study subjects. L1, L2 and L3 are depicted in Figure 3.
In Study III, the Finnish IBD cases were genotyped using the matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) analysis on a Massarray Analyzer platform (Sequenom, San Diego, CA, USA) at the Finnish Genome Center, FIMM, Helsinki, Finland. The Finnish control samples were genotyped using the Illumina 610-Quad platform as described previously (Dubois et al. 2010). Due to multiplexing incompatibility of one marker (rs2327832), it was genotyped using Taqman chemistry (Applied Biosystems, Foster city, CA, USA) in the Finnish IBD dataset. The genotyping success rate for this marker was 97.3%. In both cases and controls, the overall genotyping success rate for all the markers was between 98.2% and 99.9%. To confirm that all three genotyping platforms (Sequenom-MALDI-TOF, Illumina 610-Quad and Taqman chemistry) were consistent, some control samples were genotyped on all three platforms. No discrepancy in genotyping was found among different methods.

The Swedish UC samples and control samples were genotyped on an Illumina Hap550 array and Illumina Quad-610 Beadchips. Detailed information on samples, genotyping quality control and genotyping call rates has been provided elsewhere (McGovern et al. 2010a).

10. Power calculations

The genetic power calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html) (Purcell, Cherny & Sham 2003), using the case-control for discrete traits option, was used to calculate the power of our datasets to detect associations. In Study II (rs601338-AA (FUT2) association study), the prevalence of CelD and IBD was set to 0.01 and 0.004 respectively. The frequency of rs601338-A allele was set to 0.412 as suggested previously (McGovern et al. 2010b). For the odds ratios (ORs) 1.15-1.5, the CelD, DH and CelD + DH datasets gave 29%-99%, 9%-44% and 32%-100% power respectively. For the same ORs, the UC, CD and UC + CD datasets gave 20%-93%, 14%-77% and 26%-98% power, respectively.

In the Study III, three different allele frequencies (0.20, 0.25 and 0.30) were used to calculate the power of our study to detect association. These allele frequencies were representative of the minor allele frequencies of the CelD risk variants identified by Dubois et al. in CelD GWAS (Dubois et al. 2010). The range of ORs used for power calculations was 1.15-1.4. In the Finnish CD and UC datasets, the power to detect an association, for the mentioned ORs and allele frequencies, ranged from 18% to 57% and from 27% to 81% respectively. The power in the Swedish-Finnish UC meta-analysis was estimated to be between 55% and 99%.

11. Data analysis

For association analyses, the genotypes and affection status were imported to the BC-SNPmax LIMS database (Biocomputing platforms, Espoo, Finland) and single marker association tests and, where
relevant, two marker haplotype association tests, were performed using PLINK version v1.07 (Purcell et al. 2007). In Study III, a meta-analysis of Swedish-Finnish UC datasets was performed using the Cochran-Mantel-Haenszel function in PLINK. To check for homogeneity of ORs, the combined datasets were subjected to Breslow-Day test. No marker deviated significantly from homogeneity (P < 0.05). In the association analyses (Study II and Study III) statistical threshold for a significant association was set at P < 0.05. When multiple SNP markers are used in an analysis, the obtained P-values for association must be corrected for multiple testing (e.g. using Bonferroni correction for multiple testing). This may prove problematic in analyses where an association with a SNP has been detected using a small sample size. Multiple testing correction in such cases usually results in a loss of statistical significance (i.e. P > 0.05). Several subgroups in study III have a small sample size. In Study III, P values uncorrected for multiple testing have been reported throughout. However, in each association analysis, the option to perform 10 000 permutations was selected. This method was used as a less stringent alternative to Bonferroni correction for multiple testing. This method generates an empirical P value (P_{EMP2}) for each association and P_{EMP2} < 0.05 is considered statistically significant.
Results

1. Study I: The effect of gliadin peptides on Caco-2 cell gene transcription

The effect that the gliadin peptides have on Caco-2 cell gene transcription was studied using microarray chips and qRT-PCR method. The results of these analyses are outlined below.

1.1. Differential gene expression in Caco-2 cells after gliadin treatment

The genes on the microarrays utilized in this study are represented by probes and several genes are represented by more than one probe. Hence, the number of differentially expressed probes does not correspond to the number of differentially expressed genes. For this reason, the term probe, instead of gene, is used. A complete list of probes, which were differentially regulated (uncorrected $P < 0.01$) accompanies Publication I. A visual presentation of the data is provided in Figure 6. After six hours treatment with PT-G, the expression of 1705 probes was affected in Caco-2 cells compared to the expression in cells that were kept in medium only (MED-CTL). In contrast, 1755 probes were differentially expressed after PT treatment. The expression of 1145 probes was affected by both PT-G and PT treatment. A statistically significant (uncorrected $P < 0.05$) change in the expression of 211 probes was observed in Caco-2 cells treated with the synthetic p31-43 peptide. Out of these 211 probes, 21 probes were common between PT treatment and p31-43 treatment, whereas PT-G treatment shared ten probes with p31-43 treatment. The expression of 46 probes was affected by all three treatments compared to MED-CTL, whereas the expression of 543, 504 and 134 probes was unique to PT, PT-G and p31-43 treatments respectively.

1.2. Validation of differentially expressed genes by qRT-PCR

The observed changes in the expression of probes by different treatments were verified by qRT-PCR. Four probes corresponding to four different genes were selected based on the criteria described in the materials and methods section. Fold change (FC) values greater than one indicate up-regulation and smaller than one indicate down-regulation. In total, the expression of five genes was verified by qRT-PCR. The expression of three genes [CCAAT/enhancer binding protein alpha ($CEBP\alpha$), Kruppel-like factor 4 ($KLF4$) and sprouty-related, EVH1 domain containing 1 ($SPRED1$)] was affected by PT-G treatment. Two genes [$mitochondrial nucleoid factor 1 (MNF1)$ and $RELA$] were affected by p31-43 treatment. The qRT-PCR results for these genes are shown in Table 10.
The qRT-PCR analysis verified the change in gene expression levels observed in the microarrays for the genes CEBPa, KLF4 and SPRED1. The CEBPa gene was down-regulated to a similar degree in qRT-PCR as on the microarrays (0.66 vs. 0.71) and the level of up-regulation for KLF4 (1.78 vs. 2.0) and SPRED1 (1.50 vs. 1.66) was also comparable. For all three genes the expression levels were affected to a similar degree by negative controls, i.e. PT, PT-BSA and PT-L (Table 10). In contrast, for the p31-43 peptide affected genes, the microarray results could not be validated. The expression values in qRT-PCR for the MNF1 and RELA genes were more similar to MED-CTL and the negative control p57-68 treated cells.
2. Study II: Association of FUT2 with CelD, DH and IBD in the Finnish population

The genotype distribution in CelD (P = 0.0006) and CelD + DH (P = 0.0004) datasets deviated significantly from HWE. Although the DH, UC and CD datasets were in HWE (P > 0.05), the deviation from HWE in the UC + CD dataset was statistically significant (P = 0.034). The AA genotype of the rs601338 SNP marker, which is associated with FUT2 non-secretor status, showed significant genotype association with increased risk for CelD [P = 0.0074, OR: 1.28 (95% CI 1.05-1.56)] in the Finnish population (Table 11). In addition, rs601338-A also showed a nominal, but statistically significant, recessive association [P = 0.015, OR: 1.28 (1.05-1.56)]. The rs601338 SNP did not associate with DH subset of CelD patients under any of the tested association models (P > 0.05). Interestingly, when the association of rs601338-AA genotype was tested in the combined CelD + DH dataset, an association pattern similar to the CelD dataset alone was observed. Rs601338 showed significant genotype association of the non-secretor FUT2 genotype (rs601338-AA) in the CelD + DH datasets [P = 0.0060, OR: 1.28 (1.06-1.55)] and rs601338-A showed recessive association with disease risk in the CelD + DH dataset [P = 0.011, OR: 1.28 (1.06-1.55)]. There were approximately 18.0% non-secretors in all three patient groups, CelD, DH and CelD + DH, compared to 14.7% in control individuals.

Table 10. Validation of microarray results for selected genes by qRT-PCR.

<table>
<thead>
<tr>
<th>Gene</th>
<th>qRT-PCR</th>
<th>Microarray</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MED-CTL</td>
<td>PT</td>
</tr>
<tr>
<td>CEBPa</td>
<td>1.01</td>
<td>0.80</td>
</tr>
<tr>
<td>KLF4</td>
<td>1.06</td>
<td>1.71</td>
</tr>
<tr>
<td>SPRED1</td>
<td>1.04</td>
<td>1.48</td>
</tr>
<tr>
<td>MNF1</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>RELA</td>
<td>1.05</td>
<td></td>
</tr>
</tbody>
</table>

The qRT-PCR results for CCAAT/enhancer binding protein alpha (CEBPα), Kruppel-like factor 4 (KLF4), SPRED1: Sprouty-related, EVH1 domain containing 1 (SPRED1) are based on eight independent experiments. The qRT-PCR results for Mitochondrial nucleoid factor 1 (MNF1), REL-A: v-rel reticuloendotheliosis viral oncogene homolog A (REL-A), are based on five independent experiments. The expression of genes was normalised against Beta-2 microglobulin (B2M) gene. The values expressed for different treatments are the average expression values.
Table 11. Full-model association analysis of rs601338-A allele with risk for CeLD and IBD in the Finnish population.

<table>
<thead>
<tr>
<th>Association test</th>
<th>CTL (n=2738)</th>
<th>CeLD (n=909)</th>
<th>DH (n=116)</th>
<th>CeLD+DH (n=1025)</th>
<th>UC (n=496)</th>
<th>CD (n=280)</th>
<th>UC+CD (n=745)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HWE</td>
<td>0.81</td>
<td>0.0006</td>
<td>0.36</td>
<td>0.0004</td>
<td>0.18</td>
<td>0.084</td>
<td>0.034</td>
</tr>
<tr>
<td><strong>Genotype</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P&lt;0.05</td>
<td>0.0074</td>
<td>0.55</td>
<td>0.0060</td>
<td>0.12</td>
<td>0.18</td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>1.28</td>
<td>1.29</td>
<td>1.28</td>
<td>0.97</td>
<td>1.11</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td>(95% CI)</td>
<td>(1.05-1.56)</td>
<td>(0.79-2.09)</td>
<td>(1.06-1.55)</td>
<td>(0.74-1.28)</td>
<td>(0.78-1.56)</td>
<td>(0.82-1.30)</td>
<td></td>
</tr>
<tr>
<td><strong>Allelic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P&lt;0.05</td>
<td>0.59</td>
<td>0.62</td>
<td>0.52</td>
<td>0.12</td>
<td>0.49</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>1.03</td>
<td>1.07</td>
<td>1.04</td>
<td>0.9</td>
<td>0.94</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>(95% CI)</td>
<td>(0.92-1.15)</td>
<td>(0.82-1.40)</td>
<td>(0.93-1.15)</td>
<td>(0.78-1.03)</td>
<td>(0.78-1.12)</td>
<td>(0.82-1.04)</td>
<td></td>
</tr>
<tr>
<td><strong>Dominant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P&lt;0.05</td>
<td>0.29</td>
<td>0.97</td>
<td>0.33</td>
<td>0.044</td>
<td>0.15</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>0.92</td>
<td>0.99</td>
<td>0.93</td>
<td>0.82</td>
<td>0.83</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>(95% CI)</td>
<td>(0.79-1.07)</td>
<td>(0.68-1.46)</td>
<td>(0.80-1.08)</td>
<td>(0.68-1.00)</td>
<td>(0.65-1.07)</td>
<td>(0.71-0.99)</td>
<td></td>
</tr>
<tr>
<td><strong>Recessive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P&lt;0.05</td>
<td>0.015</td>
<td>0.31</td>
<td>0.011</td>
<td>0.83</td>
<td>0.53</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>1.28</td>
<td>1.29</td>
<td>1.28</td>
<td>0.97</td>
<td>1.11</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td>(95% CI)</td>
<td>(1.05-1.56)</td>
<td>(0.79-2.09)</td>
<td>(1.06-1.55)</td>
<td>(0.74-1.28)</td>
<td>(0.78-1.56)</td>
<td>(0.82-1.30)</td>
<td></td>
</tr>
<tr>
<td><strong>AA_NS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f(A)</td>
<td>14.7%</td>
<td>18.0%</td>
<td>18.1%</td>
<td>18.0%</td>
<td>14.3%</td>
<td>16.1%</td>
<td>15.0%</td>
</tr>
<tr>
<td></td>
<td>38.5%</td>
<td>39.2%</td>
<td>40.1%</td>
<td>39.3%</td>
<td>35.9%</td>
<td>37.0%</td>
<td>36.5%</td>
</tr>
</tbody>
</table>

HWE: Hardy-Weinberg equilibrium, OR: Odds ratio. Significant (P<0.05) P-values and their OR-values are written in bold.
CTL: Controls, AA_NS: Non-secretors with AA genotype, f(A): Frequency of allele A

In the IBD dataset, rs601338-AA showed no significant association with CD (P > 0.05), neither was there any considerable difference in the frequency of non-secretors between CD patients and controls (16.1% vs. 14.7%). However, the rs601338-AA genotype showed significant association with UC under the dominant association model [P=0.044, OR: 0.82 (0.68-1.00)], suggestive of a protective effect of the rs601338-A allele. Thus, in contrast to the rs601338-AA association with increased risk for CeLD, rs601338-G is associated with increased risk for UC. In the IBD (UC + CD) dataset, rs601338-G was associated with increased risk for IBD [P= 0.035, OR: 0.84 (0.71-0.99)]. The frequency of non-secretors in UC and IBD was 14.3% and 15%, respectively.

3. Study III: Association of CeLD risk genes with IBD and with CD/UC subphenotypes

The results of the association analyses in the different datasets are presented below. The data from the most significant findings from the association analyses are also tabulated. Detailed information for all the markers and association results is given in the original publication (publication III).
3.1. Association of CelD risk loci with CD and/or UC in the Finnish population

The SNPs which were significantly associated with CD and UC (uncorrected P < 0.05) are listed in Table 12 and Table 13 respectively. Six SNPs each were associated with CD and UC and two SNPs (rs1893217-PTPN2 and rs4819388-ICOSLG) were associated with both forms of IBD. Of the associated SNPs, the strongest association with CD was observed for SNP rs4819388-ICOSLG [P= 0.015, OR: 1.35 (1.06-1.71)] with the over-representation of the G risk allele in CD patients compared to healthy controls (81% vs. 76%). The strongest evidence for association with UC was for SNP rs242372-A [P= 0.0076, OR: 1.33 (1.08-1.64)]. In addition, seven CelD risk alleles were protective (OR < 0.90) against CD or UC and 23 risk alleles were associated with higher risk (OR > 1.10) for CD or UC (Publication III, Table 1). However, when the P-values were corrected for the number of markers tested, as described in the materials and methods section, the observed associations were no longer statistically significant (P_{EMP2} > 0.05). The association of two or three marker haplotypes with CD and/or UC was also studied, but no significant associations were found.

Table 12. CelD risk loci with significant association with CD are shown.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chr</th>
<th>Mb</th>
<th>Candidate genes</th>
<th>R_A</th>
<th>CTL</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12727642</td>
<td>1</td>
<td>7.2</td>
<td>PARK7, TNFRSF9</td>
<td>T</td>
<td>0.26</td>
<td>0.21</td>
</tr>
<tr>
<td>rs10936599</td>
<td>3</td>
<td>166.9</td>
<td>MYNN</td>
<td>A</td>
<td>0.26</td>
<td>0.31</td>
</tr>
<tr>
<td>rs13109373</td>
<td>4</td>
<td>158.8</td>
<td>FSTL4</td>
<td>T</td>
<td>0.12</td>
<td>0.16</td>
</tr>
<tr>
<td>rs653178</td>
<td>12</td>
<td>109.0</td>
<td>SH2B3, ATXN2</td>
<td>C</td>
<td>0.41</td>
<td>0.46</td>
</tr>
<tr>
<td>rs1893217</td>
<td>18</td>
<td>12.8</td>
<td>PTPN2</td>
<td>C</td>
<td>0.16</td>
<td>0.20</td>
</tr>
<tr>
<td>rs4819388</td>
<td>21</td>
<td>31.0</td>
<td>ICOSLG</td>
<td>G</td>
<td>0.76</td>
<td>0.81</td>
</tr>
</tbody>
</table>

45 SNPs were tested for their association with CD (n = 230) and 2482 controls were used in each analysis. The uncorrected P-values < 0.05 are given. The SNP positions are based on Genome build 37.1 HuRef. Mb: Genomic location in mega bases, R_A: Risk allele as in Dubois et al., CTL: Controls, RAF: Risk allele frequency.
Table 13. CelD risk loci with significant association with ulcerative colitis are shown.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chr</th>
<th>Mb</th>
<th>Candidate genes</th>
<th>R_A</th>
<th>CTL</th>
<th>UC</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs864537</td>
<td>1</td>
<td>138.7</td>
<td>CD247</td>
<td>T</td>
<td>0.70</td>
<td>0.26 0.017 1.22</td>
</tr>
<tr>
<td>rs13119723</td>
<td>4</td>
<td>118.9</td>
<td>IL2, IL21, ADAD1</td>
<td>G</td>
<td>0.10</td>
<td>0.12 0.040 1.26</td>
</tr>
<tr>
<td>rs11221332</td>
<td>11</td>
<td>124.3</td>
<td>ETS1</td>
<td>A</td>
<td>0.23</td>
<td>0.26 0.036 1.19</td>
</tr>
<tr>
<td>rs1893217</td>
<td>18</td>
<td>12.8</td>
<td>PTPN2</td>
<td>C</td>
<td>0.16</td>
<td>0.19 0.033 1.22</td>
</tr>
<tr>
<td>rs242372</td>
<td>21</td>
<td>14.3</td>
<td></td>
<td>A</td>
<td>0.12</td>
<td>0.15 0.0076 1.33</td>
</tr>
<tr>
<td>rs4819388</td>
<td>21</td>
<td>31.0</td>
<td>ICOSLG</td>
<td>G</td>
<td>0.76</td>
<td>0.20 0.016 1.24</td>
</tr>
</tbody>
</table>

45 SNPs were tested for their association with UC (n = 441) and 2482 controls were used in each analysis. The uncorrected P-values < 0.05 are given. The SNP positions are based on Genome build 37.1 HuRef. Mb: Genomic location in mega bases, R_A: Risk allele as in Dubois et al., CTL: Controls, RAF: Risk allele frequency.

3.2. Association of CelD risk loci with subphenotypes of CD and/or UC in the Finnish population

The association of CelD risk alleles with different subphenotypes of CD and UC was also investigated and the complete results are shown in publication III, Table 2. In total, 28 SNPs were associated (uncorrected P < 0.05) with CD and/or UC. Of these associations, 22 associations were specific for a CD subphenotype, 17 associations were UC subphenotype specific and 11 SNPs were associated with both CD and UC subphenotypes. The SNP rs6974491-A-ELMO1 showed strong evidence for association with early-onset UC [uncorrected P =0.0002, OR: 2.20 (1.44-3.36)]. The risk allele was present in 26% of UC patients (n=59, UC onset age: 0-18 years) but only in 14% of healthy controls. Similarly, rs2298428-T-UBE2L3 was associated with early-onset CD [uncorrected P= 5.4x10^-4, OR: 2.59 (1.61-4.19)] and the risk allele was overrepresented in CD patients compared to healthy controls (50% vs. 29%). This SNP was also associated with the non-stricturing, non-penetrating phenotype of CD [uncorrected P =0.0010, OR: 1.66 (1.22-2.25)]. All three associations remained statistically significant (P<0.05) after correcting for the multiple tests performed in the analysis.

3.3. Meta-analysis of Finnish and Swedish UC datasets

The results of the Finnish and Swedish UC meta-analysis are shown in Table 14. However, only those SNPs which were significantly associated (uncorrected P < 0.05) in either dataset (i.e. Finnish, Swedish or the combined Finnish + Swedish) are presented in the table. Seven SNPs were significantly associated with UC and three SNPs remained significantly associated (P<0.05) after correcting for the number of markers in the analysis. Two of these markers rs13003464-G-PUS10 [uncorrected P=0.0010, OR: 1.22 (1.08-1.37)] and rs4819388-G-ICOSLG [uncorrected P=0.00042, OR: 1.27 (1.11-1.45)] were associated with increased risk for UC, whereas one marker rs1738074-A-TAGAP [uncorrected P=0.0.0012, OR: 0.83 (0.73-0.93)] was protective.
The results from the Finnish-Swedish UC subphenotype meta-analysis are shown in Table 15. In total, nine SNP markers were associated (uncorrected $P < 0.05$) with one or more UC subphenotypes. The most significant findings in this analysis were strong associations with familial UC, sporadic UC and early-onset UC (onset age: 0-18 years). rs1738074-A-TAGAP was associated with familial UC [uncorrected $P=7.40 \times 10^{-5}$, OR: 0.61 (0.47-0.77) and $P_{EMP2} < 0.005$]. The associated allele showed a protective effect, being present in 43% of the controls but only in 33% of the affected individuals with a family history of UC. The SNP which showed strong evidence for association with sporadic UC was rs4819388-G-ICOSLG [uncorrected $P=0.00019$, OR: 0.75 (0.65-0.87) and $P_{EMP2} < 0.005$]. The CelD risk SNP rs6974491-A-ELMO1 was associated with early-onset UC [uncorrected $P=0.00052$, OR: 1.73 (1.26-2.37) and $P_{EMP2} < 0.05$].

Table 15. Meta-analysis of Finnish and Swedish UC datasets.

<table>
<thead>
<tr>
<th>Marker</th>
<th>R_A</th>
<th>Candidate genes</th>
<th>FIN_UC (n=441)</th>
<th>SWE_UC (n=923)</th>
<th>UC meta-analysis (n=1368)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12727642</td>
<td>A</td>
<td>PARK7, TNFRSF9</td>
<td>0.14 (0.88)</td>
<td>0.17 (0.86)</td>
<td>0.034 (0.87)</td>
</tr>
<tr>
<td>rs10903122</td>
<td>G</td>
<td>RUNX3</td>
<td>0.17 (0.90)</td>
<td>0.038 (0.83)</td>
<td>0.030 (0.88)</td>
</tr>
<tr>
<td>rs864537</td>
<td>T</td>
<td>CD247</td>
<td>0.017 (1.22)</td>
<td>0.11 (1.16)</td>
<td>0.0046 (1.19)</td>
</tr>
<tr>
<td>rs13003464</td>
<td>G</td>
<td>PUS10</td>
<td>0.064 (1.15)</td>
<td>0.015 (1.25)</td>
<td>0.0010* (1.22)</td>
</tr>
<tr>
<td>rs13119723</td>
<td>G</td>
<td>IL2, IL21</td>
<td>0.040 (1.26)</td>
<td>0.47 (0.91)</td>
<td>0.33 (1.09)</td>
</tr>
<tr>
<td>rs10806425</td>
<td>A</td>
<td>BACH2</td>
<td>0.47 (1.06)</td>
<td>0.014 (0.80)</td>
<td>0.11 (0.91)</td>
</tr>
<tr>
<td>rs1738074</td>
<td>A</td>
<td>TAGAP</td>
<td>0.059 (0.87)</td>
<td>0.021 (0.81)</td>
<td>0.0012* (0.83)</td>
</tr>
<tr>
<td>rs653178</td>
<td>G</td>
<td>SH2B3, ATXN2</td>
<td>0.066 (1.15)</td>
<td>0.082 (1.17)</td>
<td>0.016 (1.15)</td>
</tr>
<tr>
<td>rs1893217</td>
<td>G</td>
<td>PTPN2</td>
<td>0.033 (1.22)</td>
<td>0.21 (0.87)</td>
<td>0.43 (1.06)</td>
</tr>
<tr>
<td>rs242372¹</td>
<td>A</td>
<td></td>
<td>0.0076 (1.33)</td>
<td>0.19 (0.85)</td>
<td>0.20 (1.11)</td>
</tr>
<tr>
<td>rs4819388</td>
<td>G</td>
<td>ICOSLG</td>
<td>0.016 (1.24)</td>
<td>0.0086 (1.30)</td>
<td>0.00042* (1.27)</td>
</tr>
</tbody>
</table>

The number of controls in the Finnish- and Swedish- UC meta-analysis was 2482 and 341 respectively. The SNPs and $P$-values showing more significant association in the UC meta-analysis compared to the individual analysis (FIN_UC and/or SWE_UC) are underlined. The associations which remained significant after multiple testing correction are marked with asterisk (*) ($P_{EMP2}<0.05$). Genotypes for only 2171 controls were available. R_A: Risk allele as in Dubois et al.
Table 15. UC-subphenotype meta-analysis of Finnish- and Swedish UC datasets.

<table>
<thead>
<tr>
<th>Marker</th>
<th>R_A</th>
<th>Candidate genes</th>
<th>FAMILIAL UC</th>
<th>SPORADIC UC</th>
<th>UC ONSET AGE: 0-18yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FIN (n=115)</td>
<td>SWE (n=51)</td>
<td>FIN+SWE (n=166)</td>
<td>FIN (n=326)</td>
</tr>
<tr>
<td>rs10903122</td>
<td>G</td>
<td>RUNX3</td>
<td>0.17 (0.89)</td>
<td>0.087 (0.84)</td>
<td>0.029 (0.87)</td>
</tr>
<tr>
<td>rs864537</td>
<td>T</td>
<td>CD247</td>
<td>0.015 (1.26)</td>
<td>0.064 (1.21)</td>
<td>0.0022 (1.24)</td>
</tr>
<tr>
<td>rs859637</td>
<td>A</td>
<td>FASLG, TNFSF18</td>
<td>0.070 (0.78)</td>
<td>0.20 (0.76)</td>
<td>0.03 (0.77)</td>
</tr>
<tr>
<td>rs2816316</td>
<td>A</td>
<td>RGS1</td>
<td></td>
<td></td>
<td>0.096 (0.68)</td>
</tr>
<tr>
<td>rs13003464</td>
<td>G</td>
<td>PUS10</td>
<td>0.076 (1.16)</td>
<td>0.038 (1.24)</td>
<td>0.0021 (1.23)</td>
</tr>
<tr>
<td>rs917997</td>
<td>A</td>
<td>IL18RAP, IL18R1</td>
<td>0.88 (1.03)</td>
<td>0.043 (0.57)</td>
<td>0.17 (0.81)</td>
</tr>
<tr>
<td>rs10936599</td>
<td>A</td>
<td>MYNN</td>
<td></td>
<td></td>
<td>0.16 (1.32)</td>
</tr>
<tr>
<td>rs10806425</td>
<td>A</td>
<td>BACH2</td>
<td>0.09 (1.17)</td>
<td>0.0043 (0.75)</td>
<td>0.19 (0.91)</td>
</tr>
<tr>
<td>rs1738074</td>
<td>T</td>
<td>TAGAP</td>
<td>0.0012 (0.63)</td>
<td>0.039 (0.64)</td>
<td>0.000074** (0.61)</td>
</tr>
<tr>
<td>rs6974491</td>
<td>A</td>
<td>ELMO1</td>
<td></td>
<td></td>
<td>0.00020 (2.20)</td>
</tr>
<tr>
<td>rs2762051</td>
<td>A</td>
<td></td>
<td>0.094 (1.30)</td>
<td>0.024 (0.50)</td>
<td>0.86 (1.03)</td>
</tr>
<tr>
<td>rs4819388</td>
<td>G</td>
<td>ICOSLG</td>
<td></td>
<td></td>
<td>0.0022 (1.38)</td>
</tr>
</tbody>
</table>

The number of controls in the FIN+SWE UC subphenotype meta-analysis was 2823. The associations which remained significant after multiple testing correction are marked with asterisk (*) (* P<0.05 and ** P<0.005). R_A: Risk allele as in Dubois et al.
Discussion

The main purpose of this study was to explore the shared genetic background in CelD and IBD and investigate the impact and interplay of selected genes and environmental factors on CelD and IBD susceptibility. In CelD, gluten peptides are the disease-causing environmental triggers and critical epitopes for T-cell driven pathogenesis. However, a complete picture of how (and if) gluten affects epithelial cell gene transcription remains elusive (i.e. environmental effects on genes). The gut microbiota is known to affect the development of the innate and adaptive immune system and polymorphisms in the \textit{FUT2} gene have been shown to affect the composition of gut microbiota (genetic influence on environment). Inflammation in chronic inflammatory diseases (such as CelD and IBD) is sustained by effector cells of the immune system. Inflammation is the end product of an intricate signaling cascade which is affected by changes in the transcription and/or translation pattern of multiple genes. Clinically distinct inflammatory diseases are likely to share underlying disease risk due to similar changes in some of these genes.

1. Modest effect of gliadin peptides and novel, non-enzymatic effects of pepsin and trypsin on Caco-2 gene transcription

The effects of gliadin peptides on epithelial cell gene transcription were investigated using gene expression microarrays and select genes from this assay were verified using qRT-PCR. The qRT-PCR assay validated that the expression of three genes, \textit{CEBPa}, \textit{KLF4} and \textit{SPRED1} is indeed affected by PT-G. However, these effects are unlikely to be direct effects of gliadin peptides as similar changes in gene expression (with respect to direction and magnitude) were also brought about by the negative controls PT, PT-BSA and PT-L. Furthermore, the expression values of genes affected by the cytotoxic p31-43 peptide were similar to those of the control, suggesting that these genes may be false positives. As compared to the MED-CTL, after exposure to PT-G a unique set of 504 probes was differentially expressed in Caco-2 cells and only three genes were verified by qRT-PCR. It is practically unrealistic to replicate all the affected genes by qRT-PCR. Thus it cannot be concluded that gliadin peptides have no effect on enterocyte gene expression. The possible conclusions that can be drawn from the data are as follows: the direct effects of gliadin derived peptides on epithelial cell gene transcription are marginal; pepsin- and trypsin- derived peptides are able to affect enterocyte gene expression and the effect of the gliadin peptides on enterocytes requires the presence of other components of the immune system (i.e. effector cells and/or cytokines) or mucosal environment.

Several previous studies (Table 16) have attempted to identify genes that are differentially regulated in CelD (Diosdado et al. 2004, Juuti-Uusitalo et al. 2004, Simula et al. 2010, Bracken et al. 2008). This
study is not directly comparable to those studies as they have been performed on biopsy samples obtained from CeID patients or on enterocytes isolated from CeID biopsies. Although no gene expression microarray study investigating the effects of gliadin peptides on cultured epithelial cells has been performed, many comparisons can be made between this study and previous microarray studies.

Juuti-Uusitalo et al. compared gene expression in duodenal biopsies from untreated- and treated-CeID patients with that of healthy controls (Juuti-Uusitalo et al. 2004). There were 156 and 60 genes differentially regulated in untreated and treated groups respectively compared to controls. Ninety-eight genes were differentially regulated between the untreated and treated groups. Nine differentially expressed genes reported by Juuti-Uusitalo et al. were also among the differentially regulated genes in this study (Table 16). However, three of the nine genes were affected by both PT and PT-G in our study and two genes were affected by PT only. The \textit{RALY} (RNA-binding protein, autoantigenic) gene was up-regulated in Caco-2 cells exposed to p31-43 peptide and was also over-expressed in treated CeID biopsy samples. The function of \textit{RALY} gene is however poorly understood.

The genes, \textit{CD59 antigen} (\textit{CD59}), \textit{Ephrin B2} (\textit{EFNB2}) and \textit{Myosin VI} (\textit{MYO6}) genes were affected by PT-G. The difference in FC was almost identical for \textit{RALY}, \textit{EFNB2} and \textit{MYO6} between the two studies and all were up-regulated. It is of note that FC for these genes in both studies are modest (i.e. FC ≤ 1.60). The \textit{CD59} gene was 1.5 fold down-regulated in the study by Juuti-Uusitalo et al. whereas it was 1.4 fold up-regulated in our study. The \textit{CD59} gene encodes a glycosylphosphatidylinositol (GPI)-linked protein which is the cell surface inhibitor of the membrane attach complex on human cells (Piccoli et al. 2011). The CD59 protein is expressed in most tissues and in all circulating cells, including erythrocytes and leukocytes. Interestingly, the expression of CD59 on cell surface also makes the cells more susceptible to NK cell-mediated cytotoxicity (Marcenaro et al. 2003, Omidvar et al. 2006) and genetic variants within this gene have been linked to CeID in a previous study (Vidal et al. 2009). It is possible that under physiological conditions (e.g. biopsy samples), the \textit{CD59} gene expression is affected by other factors (other cells, cytokines etc.) also which are absent \textit{in vitro} assays (cultured cells). This is likely to be the reason for differential expression pattern of this gene between the two studies. The \textit{EFNB2} gene was over-expressed in both untreated CeID biopsies and Caco-2 cells stimulated with PT-G. This gene is important in T cell development and immune regulation (Yu et al. 2003, Luo et al. 2011). Similarly, \textit{MYO6} gene was over-expressed in both studies and has been reported to facilitate the translocation of endocytic vesicles from cell peripheries (Aschenbrenner, Lee & Hasson 2003).
In another study (Diosdado et al. 2004), the differences in gene expression were studied in biopsy specimens obtained from CelD patients on a GFD and CelD patients on a gluten containing diet. Four genes affected by PT and PT-G in this study were reported to be differentially expressed. Bracken et al. examined the gene expression profile of intestinal epithelial cells using microarrays (Bracken et al. 2008). However, the difference between our study and the one by Bracken et al. is that the epithelial cells in the latter study were isolated from biopsy specimens from CelD patients with active disease and from control subjects. In addition, the isolated enterocytes were not exposed to PT-G. As shown in Table 16 two genes, \textit{ETS-domain protein (ELK4)} which is also known as \textit{SRF accessory protein 1 (SAP1)} and \textit{prion protein (PRNP)}, were affected by PT-G only in this study. The \textit{ELK4} gene was approx. 1.4 fold up-regulated in both studies whereas the \textit{PRNP} gene was 1.3 fold up-regulated in our study and approximately two fold down-regulated in the other study. The \textit{ELK4} gene is required for thymocyte positive selection (Costello et al. 2004) whereas the \textit{PRNP} gene is over-expressed in response to endoplasmic reticulum (ER) stress and delays ER stress-induced cell death. The fact that these genes were affected by PT-G only (and not PT), and given their role in T cell development, immune regulation, apoptosis and endocytic vesicle translocation, further studies are warranted to ascertain their role in CelD pathogenesis.

Pepsin and trypsin are proteolytic enzymes with different substrate specificities which participate in the digestion of food by breaking down food protein into peptides. Interestingly, in addition to their role in food digestion, effects unrelated to the digestive process have been reported. Pepsin is released as a pro-form zymogen, pepsinogen, by chief cells in the stomach. Upon activation under acidic conditions in the stomach, pepsinogen cleaves itself in an autocatalytic fashion to make pepsin. Pepsin displays maximum activity at pH 2.0 and is inactive at pH 6.5 or higher, however it remains stable until pH 8.0 and can be reactivated by lowering the pH.(Johnston et al. 2007).
Table 16. Comparison of microarray data with previous studies

<table>
<thead>
<tr>
<th>Microarrays</th>
<th>Previous studies</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLX4</td>
<td>1.60</td>
<td>1.50</td>
</tr>
<tr>
<td>DLX4</td>
<td>1.60</td>
<td>1.50</td>
</tr>
<tr>
<td>DLX4</td>
<td>1.60</td>
<td>1.50</td>
</tr>
<tr>
<td>JUNB</td>
<td>3.40</td>
<td>2.90</td>
</tr>
<tr>
<td>NAB2</td>
<td>2.00</td>
<td>1.90</td>
</tr>
<tr>
<td>PROCX2</td>
<td>1.40</td>
<td>1.43</td>
</tr>
<tr>
<td>SLC25A6</td>
<td>1.80</td>
<td>1.70</td>
</tr>
<tr>
<td>SLC25A6</td>
<td>1.80</td>
<td>1.70</td>
</tr>
<tr>
<td>RALY</td>
<td>1.60</td>
<td>1.43</td>
</tr>
<tr>
<td>CD59</td>
<td>1.40</td>
<td>0.66</td>
</tr>
<tr>
<td>EFNB2</td>
<td>1.40</td>
<td>1.41</td>
</tr>
<tr>
<td>MYO6</td>
<td>1.27</td>
<td>1.27</td>
</tr>
<tr>
<td>HSPA6</td>
<td>2.60</td>
<td>2.90</td>
</tr>
<tr>
<td>MSII2</td>
<td>0.75</td>
<td>0.72</td>
</tr>
<tr>
<td>POLD3</td>
<td>0.66</td>
<td>0.61</td>
</tr>
<tr>
<td>RGS16</td>
<td>2.70</td>
<td>2.57</td>
</tr>
<tr>
<td>ELK4</td>
<td>1.37</td>
<td>1.45</td>
</tr>
<tr>
<td>UBE3A</td>
<td>0.82</td>
<td>0.77</td>
</tr>
<tr>
<td>NUDT2</td>
<td>0.72</td>
<td>0.74</td>
</tr>
<tr>
<td>PRNP</td>
<td>1.31</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Expression values > 1 indicate up-regulation and < 1 indicate down-regulation. Condition refers to the experimental condition used in the original study. UCD: Biopsy from untreated CD, TCD: Biopsy from treated CD, HC: Biopsy from healthy control, MIII-G+: Biopsy from Marsh-III stage CD patient on gluten containing diet, MIII-G-: Biopsy from Marsh-III stage CD patient on gluten-free diet, Ent. CD: Epithelial cells isolated from duodenal biopsies of CD patients, Ent. HC: Epithelial cells isolated from duodenal biopsies of healthy controls. *In the original study, the expression values were given for MIII-G-/G- and comparable values for MIII-G+/G- were obtained (1/MIII-G-/G-) in the study.

Bathoorn et al. reported that, in a pH- and concentration-dependent manner, pepsin has cytotoxic effects on bronchial epithelial cells and also promotes inflammation by inducing the production of inflammatory cytokines (IL-8 and IL-6) (Bathoorn et al. 2011). These effects were more pronounced at lower pH (at pH 1.5 compared to pH 2.5). Furthermore, the effect of pepsin (at pH 7.0) on normal human laryngeal primary epithelial cells and hypopharyngeal FaDu squamous cell carcinoma (SCC)
cells has also been studied (Johnston et al. 2012). It was reported that the expression of 27 genes in FaDu cells and 31 genes in laryngeal primary epithelial cells is differentially regulated after exposure to pepsin at neutral pH. The treatment with pepsin also promoted proliferation in these cells. Trypsin is a serine protease which is produced in the pancreas as an inactive zymogen, trypsinogen. It is secreted via the pancreatic duct in the duodenum where it is cleaved and activated under non-acidic conditions (pH-7.5-8.0) by enteropeptidase (Kunitz 1939). Trypsin is a natural activator of proteinase-activated receptor-2 (PAR2) which belongs to a family of G protein-coupled receptors (GPCRs). In the GI, PAR2 is involved in the regulation of exocrine secretion, modulation of smooth muscle motility and plays an important role in mucosal inflammation throughout the GI (Kawabata, Matsunami & Sekiguchi 2008, Tanaka et al. 2008). Furthermore, a study in mice showed that trypsin inhibited lipopolysaccharide (LPS) signaling by degrading TLR4 accessory molecules independent of PAR2 activation (Komatsu et al. 2012). In a microarray study, the effect of PAR2 activation by trypsin on human embryonic kidney 293 (HEK293) cell gene expression was investigated (Suen et al. 2010). The expression of several genes involved in cellular metabolism, cell cycle, inflammation and the mitogen-activated protein kinase (MAPK) pathway was affected. Among the genes affected by pepsin and trypsin treatments in these studies (Johnston et al. 2012, Suen et al. 2010) five pepsin-affected genes and 16 trypsin-affected genes were also differentially expressed in this study (Table 17). For all, but two trypsin-affected genes, these changes were in the same direction and of comparable magnitude. Of the five pepsin-affected genes, one gene (BRCA1) was down-regulated in both studies, one gene (ETS2) was up-regulated in FaDu and Caco-2 cell, but not in primary epithelial cells. The remaining three genes were down-regulated in the previous study but up-regulated in this study. The difference in the direction of change for some genes in the studies may depend upon different origin of the cells being compared. Nonetheless, it is interesting that trypsin affects gene transcription.

Taken together, data from our study, the three CelD studies (Diosdado et al. 2004, Juuti-Uusitalo et al. 2004, Bracken et al. 2008) and the gene expression studies using pepsin and trypsin (Johnston et al. 2012, Suen et al. 2010) suggest that small changes in gene expression in response to gliadin peptides are possible. It is evident that pepsin and trypsin affect gene transcription in vitro and under physiological conditions as the CelD biopsies were not exposed to exogenous pepsin and trypsin. Given that pepsin and trypsin were inactivated by heating, the in vitro effects on gene expression may be mediated by non-enzymatic epitopes of these proteases. Such effects have not been reported earlier and further investigations into the non-enzymatic effects of pepsin and trypsin are warranted.
Table 17. The pepsin and trypsin affected genes which are also affected by PT and/or PT-G are shown

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Trypsin</th>
<th>Pepsin</th>
<th>PT-G</th>
<th>PT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene/Cell type</td>
<td>HEK 293</td>
<td>FaDu</td>
<td>Primary</td>
<td>Caco-2</td>
</tr>
<tr>
<td>CDT1</td>
<td>0.8</td>
<td></td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>DUSP1</td>
<td>1.3</td>
<td>2.3</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>DUSP10</td>
<td>0.9</td>
<td>1.3</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>FOSL1</td>
<td>2</td>
<td>3.2</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>ING4</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>KLF5</td>
<td>1.6</td>
<td>1.8</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>PTGS2</td>
<td>3.4</td>
<td>2</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>PTPN14</td>
<td>1.4</td>
<td>1.3</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>TSC22D3</td>
<td>0.4</td>
<td>1.3</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>TXNIP</td>
<td>0.5</td>
<td>0.4</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>GADD45A</td>
<td>1.4</td>
<td></td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>GADD45G</td>
<td>0.7</td>
<td></td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>GTF2H1</td>
<td>1.1</td>
<td></td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>NAB1</td>
<td>1.4</td>
<td>1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAB2</td>
<td>2.1</td>
<td></td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>KLF5</td>
<td>1.6</td>
<td></td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>TERT</td>
<td>0.5</td>
<td></td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>ETS2</td>
<td>1.9</td>
<td>0.6</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>BRCA1</td>
<td>0.3</td>
<td>0.5</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>BAD</td>
<td>0.4</td>
<td>0.6</td>
<td></td>
<td>1.3</td>
</tr>
<tr>
<td>BAX</td>
<td>0.5</td>
<td></td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>CDKN1A</td>
<td>0.6</td>
<td></td>
<td>2.4</td>
<td>2.6</td>
</tr>
</tbody>
</table>

The data for trypsin and pepsin affected genes is obtained from previously published studies. The numbers represent fold change after respective treatment. HEK 293: Human embryonic kidney 293 cells, FaDu: squamous cell carcinoma cells (epithelial), Primary: laryngeal primary epithelial cells, Caco-2: human colon carcinoma cells.

Gliadin peptides have been reported to exert multiple effects on enterocytes. The most prominent and consistently reported effects are apoptosis, oxidative stress and proliferation. In a study by Giovannini et al., Caco-2 cells were exposed to PT-G, and an increased rate of apoptosis (30%) compared to the control cells was reported (Giovannini et al. 2000). In a follow-up study by the same group, it was suggested that enterocyte apoptosis induced by PT-G is through the Fas-FasL pathway.
and can be inhibited using caspase-3 and caspase-8 inhibitors (Giovannini et al. 2003). Furthermore, incubation of Caco-2 cells with PT-G for six hours resulted in increased expression of the Fas and FasL genes. This study is similar to our study in experimental design, however, unlike Giovannini et al., no change in Fas/FasL gene expression after stimulation with PT-G or p31-43 peptide was observed in our study. A crucial difference between our study and the one by Giovannini et al. is the use of blank PT control in our study. It enabled us to filter out background genes, i.e. those affected by PT. Although Fas/FasL gene expression did not change in our experiments, the expression of the Fas-activated serine/threonine kinase (FASTK) gene, and the caspase 8 (CASP8) gene was down-regulated in cells exposed to PT. The FASTK gene is rapidly activated in Fas-mediated apoptosis (Tian et al. 1995). Another pro-apoptosis gene, FAST kinase domain containing protein 2 (FASTKD2), was down-regulated by both PT and PT-G treatment. These data suggest that PT and PT-G exert anti-apoptotic effect on epithelial cells. In a similar experimental setup, Sakly et al. attempted to unravel the mechanism of gliadin toxicity to Caco-2 cells (Sakly et al. 2006). They found growth inhibitory effects of gliadin peptides on enterocytes which were not due to increased apoptosis or necrosis. As discussed in section 3.6, it is possible that increased apoptosis in enterocytes is a result of an interaction between FasL expressing IELs and Fas expressing enterocytes (Moss et al. 1996, Maiuri et al. 2001, Ciccocioppo et al. 2001). IL-15 has been shown to induce Fas expression on enterocytes (Maiuri et al. 2000) and apoptosis induced by gliadin and IL-15 can be abolished by anti-IL-15 antibodies. These studies implicate the Fas-FasL pathway in enterocyte apoptosis which also requires the activation of LPMNCs to produce IL-15. Whether gliadin peptides can directly induce enterocyte apoptosis after six hours, independent of LPMNCs, remains controversial.

IL-15 has also been implicated in enterocyte proliferation (Barone et al. 2011). Overnight stimulation of Caco-2 cells with gliadin peptides induces IL-15 gene transcription, which is highly dependent on pre-existing IL-15 in the cell (Barone et al. 2011). Shorter stimulations (three hours or six hours) did not result in an increase in IL-15 gene transcription. Our results are in line with this study as no change in the expression of IL-15 was observed after six hours, suggesting that a longer exposure time is required for this effect to take place. The gliadin p31-43 peptide was shown to accumulate in lysosomes for a prolonged period and induce cellular oxidative stress which eventually leads to the initiation of inflammatory responses (Luciani et al. 2009). Oxidative stress is known to activate the NFkB pathway and increased NFkB activity has been reported in CelD mucosa (Maiuri et al. 2003b, van den Berg et al. 2001). The observation that IL-15 production in rhinovirus infected macrophages is dependent on NFkB activation (Laza-Stanca et al. 2011) connects gliadin induced oxidative stress, IL-15 and proliferation. It is conceivable that in epithelial cells, gliadin induces oxidative stress
followed by NFκB activation and IL-15 gene transcription. Given the role of IL-15 in enterocyte apoptosis and proliferation, this is an attractive hypothesis which warrants further studies.

2. Association of rs601338-A-FUT2 with CelD, DH and IBD in the Finnish population

The FUT2 gene encodes α-1,2-fucosyltransferase enzyme (Figure 5) which controls the expression of A-, B- and H- blood group antigens in mucus and other body secretions such as saliva and breast milk (Combs 2009). The secreted and expressed antigens in the gut control the composition of microbiota which in turn protects the host from pathogenic microbes and also affects the development of the mucosal immune system (Cieza et al. 2012, Shirato 2011, Makivuoikko et al. 2012, Maynard et al. 2012, Boren et al. 1993). The FUT2 rs601338-AA genotype was associated with CelD in the Finnish population. Consistent with the role of FUT2 in affecting the composition of intestinal microbiota (Rausch et al. 2011), several studies have reported a shift in gut microbial communities in CelD patients compared to healthy individuals. Schippa et al. studied the mucosa-associated microbiota in a group of children with CelD and healthy controls (Schippa et al. 2010). Highly significant differences were found in the prevalence of B. vulgatus (85%) and E. coli (95%) in CelD patients compared to controls (20% in both cases). Differences in the gut microbial community between adult and pediatric CelD patients compared to each other and also compared to healthy controls have been reported (Nistal et al. 2011, Pozo-Rubio et al. 2012, Sanz, De Pama & Laparra 2011). Furthermore, changes in the intestinal microbial community have been shown to modulate mucosal immunity. De Palma et al. studied the effect of enterobacteria (E. coli CBL2 and Shigella CBD8) isolated from CelD patients, two strains of bifidobacteria and gliadin on phenotypic and functional features of monocyte derived dendritic cells (MDDCs) (De Palma et al. 2012). The results from this study showed that enterobacteria induced marked changes in MDDC morphology and induced an increase in the production of inflammatory cytokines (IFN-γ, TNF-α and IL-12). In contrast, these changes were seen to a much lesser extent in bifidobacteria exposed cells. The bifidobacteria also reduced gliadin-induced IFN-γ production and increased IL-10 secretion by MDDCs. Similarly, studies in rats kept under germ-free conditions show that E. coli CBL2 in combination with IFN-γ and gliadin leads to a reduction in the number of mucus-secreting goblet cells in the intestine (Cinova et al. 2011). Taken together, these studies and the association of FUT2 non-secretor status with CelD suggest that FUT2 indirectly affects the risk for CelD by modulating gut microbiota and the immune system. An optimal composition of the gut microbiota may be particularly important in the childhood as multiple infections in early life are associated with increased risk for CelD (Myleus et al. 2012). The use of antibiotics affects the gut microbiota. However, in a study by Myleus et al. no
significant association was found between antibiotic treatment in early childhood and increased risk for CelD (Myleus et al. 2012).

In contrast to the association of FUT2 non-secretor status (rs601338-AA) with CelD, the association of rs601338-FUT2 with DH was not statistically significant (P > 0.05). DH is an extraintestinal manifestation of CelD and some patients do not experience gastrointestinal symptoms. Wacklin et al. investigated if differences in the composition of intestinal microbiota lead to different clinical manifestations of CelD, i.e. GI symptoms, DH, anemia or mixed symptoms (Wacklin et al. 2013). This study reported differences in mucosal microbial diversity and composition in CelD patients with GI symptoms and anemia compared to healthy controls. However, the microbial composition and diversity in DH patients was similar to that of controls. It is of note that in our study both, the frequency of non-secretors in the DH and CelD group (DH: 18.1% and CelD: 18%) and the effect sizes (OR) in both groups (DH: 1.29 and CelD: 1.28) were very similar. This suggests that the non-secretor status may be associated with DH also. It is likely that the lack of association with DH in our study may be due to lack of power in the DH group owing to fewer patients. This is further supported by the association of the rs601338-AA genotype of the FUT2 gene with the combined CelD + DH group which had several times more patients than the DH group alone. The association in the CelD + DH group showed stronger statistical significance (P=0.006), whereas the effect size (OR: 1.28) remained comparable. But given that the gut microbiota richness and diversity between CelD patients with GI symptoms and those with DH was similar, the role of the FUT2 polymorphisms in DH is unclear.

Rausch et al. conducted a study where they investigated the composition of gut microbiota in CD patients and healthy subjects conditioned on FUT2 secretor and non-secretor status (Rausch et al. 2011). There was a general change in the composition of the gut-associated microbial community in CD patients compared to healthy subjects. In addition, there were also significant secretor-status-dependent differences in microbiota composition providing some evidence that FUT2 affects the composition of gut microbiota and may even play a role in IBD pathogenesis. There are differences in the composition of gut microbiota in individuals with CD and UC compared to healthy individuals and also between both disease groups (Andoh et al. 2011, Sokol et al. 2006, Dicksved et al. 2008, Willing et al. 2009). The association of rs601338-AA with CD and a lack of association of this SNP with UC has been reported in another population of European descent (McGovern et al. 2010b). However, we did not find any evidence for association of the FUT2 non-secretor genotype (rs601338-AA) with CD. The dissimilarity between our results and those obtained by McGovern et al. may depend upon differences in the studied populations, in the microbial species involved or on genetics. The CD susceptibility gene NOD2 has been shown to affect the composition of gut
microbiota, both in mice and humans (Li et al. 2012, Rehman et al. 2011, Petnicki-Ocwieja et al. 2009) and several population-specific mutations in the NOD2 gene in the Finnish population have been identified (Lappalainen et al. 2008). It is plausible that different gene/genes affects/affect the gut microbial diversity and richness in the Finnish population. It is also conceivable that the gut microbiota at steady-state is a result of an interaction between products of different genetic variants of multiple genes. In addition, the lack of association may also be attributed to fewer patients in the CD group. Interestingly, there was some evidence for the association of rs601338-AA genotype (non-secretor status) with protection (OR: 0.82) from UC. Under the dominant association model, this suggests that the rs601338-A allele confers protection against UC. A similar mode of association was evident in the combined UC + CD (OR: 0.84) dataset. The secretor status, defined by the FUT2 A385T variant, has been associated with increased risk for CD in the Japanese population (Miyoshi et al. 2011), but a similar association with UC has not been reported in any population. In the study be McGovern et al., the FUT2 non-secretor status was associated with increased risk for CD but no association with UC was found (McGovern et al. 2010b). Our results are in stark contrast to these findings. The reason for differential association of the non-secretor status (rs601338-AA) with UC and CD between the two studies is most likely due to genetic differences between the studied populations. It is also possible that there are multiple common factors which influence the risk for UC or CD. Enrichment of certain factors in a population may confer increased risk for UC or CD in that population. Thus, having a certain genotype at one locus may not be a decisive factor in disease susceptibility. Association studies in different populations are required to further understand the role of secretor/non-secretor status in IBD susceptibility.

3. Several CelD susceptibility genes also affect the risk for IBD.

The association of some CelD risk loci with IBD in the Finnish population suggests some overlap in genetic predisposition. The genetic regions harboring the ICOSLG and PTPN2 genes were associated with both CD and UC in the Finnish population, but did not reach statistical significance after correcting for the number of tests (i.e. \( P_{\text{EMP2}} > 0.05 \)). The product of the ICOSLG gene is expressed on APCs, B cells and endothelial cells and it serves as a ligand for inducible costimulator (ICOS) molecule, expressed on T cells. When a MHC bound foreign antigen is presented by APCs to antigen-specific T cells, an interaction between ICOS and ICOSLG is crucial for the proper activation of T cells in response to the antigen. The association of ICOSLG with CD and UC has been demonstrated in other populations previously (Anderson et al. 2011, Franke et al. 2010). The lack of statistical significance for this association with CD and UC in our dataset is likely due to small sample size. The rs1893217-PTPN2 region was also nominally associated with CD location (left or right colitis) and
type (penetrating and fistulising) and with pediatric UC and with UC extensive colitis (Publication III, Table 2). An inappropriate immune response to gut-resident microbes has been implicated in mucosal inflammation in CD and UC. Recently, the PTPN2 gene was shown to be important in the maintenance of the gut epithelial barrier and the modulation of innate immune responses to bacterial antigens (McCole 2012, Scharl et al. 2012). Furthermore, genetic variants within the PTPN2 gene have been associated with susceptibility to both UC and CD in the German population (Glas et al. 2012). In addition, a SNP, different from the one in our study, in the PTPN2 harboring genomic region was associated with the stricturing phenotype in CD. Interestingly, an epistatic interaction was observed in UC between a SNP tagging the CD gene ATG16L1 and a PTPN2-tagging SNP. In a population-specific manner, a genetic variant (rs2542151-G) in the PTPN2 region has been associated with increased risk for UC in the Chinese population, whereas the other allele of the same marker (rs2542151-A) was associated with increased risk for CD in the Italian population (Latiano et al. 2011). These studies suggest a role for PTPN2 in both UC and CD susceptibility but in a population-specific manner. The association of rs1893217-PTPN2 in the Finnish population may be a true association although it did not reach statistical significance (P $\text{EMP}_2 > 0.05$) due to a smaller sample size. Similarly the ICOSLG gene tagging SNP was associated with multiple subphenotypes of both CD and UC, but did not reach statistical significance (P $\text{EMP}_2 > 0.05$) after multiple testing correction. These associations are thus interesting, but require validation in a larger cohort of Finnish IBD patients. In the UC subphenotype meta-analysis of Finnish and Swedish datasets, rs4819388-ICOSLG was significantly associated with sporadic UC (Table 15) and this association remained significant (P $\text{EMP}_2 < 0.05$) after correcting for the number of markers used in the analysis.

A SNP (rs6974491-ELMO1) on chromosome seven showed strong association with pediatric UC (OR: 2.20, P $\text{EMP}_2 < 0.05$) and was also nominally associated with left-sided colitis in UC in the Finnish population (Publication III, Table 2). Furthermore, rs6974491-ELMO1 was also associated with pediatric UC (OR: 1.73, P $\text{EMP}_2 < 0.05$) in the Finnish-Swedish UC subphenotype meta-analysis. ELMO1 is involved in the phagocytosis of apoptotic cells and changes in the cell shape (van Ham, Kokel & Peterson 2012, Elliott et al. 2010). It is already well established that genes involved in autophagy (NOD2, IRGM, ATG16L1) contribute to the pathogenesis of CD (Rioux et al. 2007, Cooney et al. 2010). Furthermore, in an association study of IRGM variants with UC in the German population, Glas et al. also found weak, but statistically significant, association of several IRGM haplotypes with UC susceptibility (Glas et al. 2013). The association of ELMO1 with UC is interesting because studies on microbiota composition in UC patients report an overrepresentation of pathogenic bacteria (Escherichia/Shigella) in the gut (Lepage et al. 2011) and Shigella has been shown to use the ELMO1-Dock180 complex for entry into epithelial cells (Handa et al. 2007). It is possible that certain ELMO1
genetic variants promote infection by *Shigella* leading to epithelial damage and activation of the immune system. The delineation of this mechanism requires further studies. Nevertheless, association of *IRGM* and *ELMO1* variants with UC suggest that autophagy may also contribute to UC pathogenesis to some extent.

The rs2298428 marker SNP lies within the *YDJC* gene on chromosome 22 and causes a missense mutation in this gene. Although this marker did not associate with CD/UC in the Finnish population, in the Finnish CD subphenotype meta-analysis, it was strongly (OR: 2.59, \(P_{EMP2} < 0.005\)) associated with pediatric CD (Publication III, Table 2). A less significant association (OR: 1.66, \(P_{EMP2} < 0.05\)) was also observed with the non-stricturing and non-penetrating phenotype of CD. Distinct genetic variants in this region are associated with different immunological diseases, such as rheumatoid arthritis (Zhernakova et al. 2011), cutaneous systemic sclerosis (Hasebe et al. 2012) and SLE (Wang et al. 2012). Despite association of this region with multiple diseases, the role of the *YDJC* gene remains poorly understood in disease pathogenesis. The rs2298428 marker lies in the vicinity of the *UBE2L3* gene. Interestingly, in a study by Fransen et al. (Fransen et al. 2010) the SNP rs2298428-T-*UBE2L3* was associated with higher expression levels of the *UBE2L3* gene in the Dutch population. However, in another study investigating common susceptibility loci between CelD and CD, genetic variants within the *UBE2L3* gene were not associated with CD in the Dutch population (Festen et al. 2011). It should however be noted that, in the Finnish IBD dataset, rs2298428-T did not associate with CD or UC but only with specific subphenotypes of these diseases. A lack of association in the whole CD dataset in our study and that by Festen *et al*. suggests that this SNP confers susceptibility in a phenotype-specific manner. The *UBE2L3* gene encodes an ubiquitin-conjugating enzyme and its functional role in IBD is not clear. Nonetheless, it may be an interesting risk gene in IBD as it may have a role in modulating the NFkB pathway as suggested by Fransen *et al*. (Fransen et al. 2010).

Two SNPs (rs9295089 and rs1738074) in the genomic region harboring the *TAGAP* gene were investigated for their association with IBD. The *TAGAP* gene encodes T-cell activation GTPase-activating protein, but the functional role of this protein has not been confirmed. The two SNPs in the genomic locus harboring the *TAGAP* gene are located close to each other on chromosome six and are in LD (\(D' = 1, r^2 = 0.19\)). The SNP marker rs1738074 was found to yield the strongest association signal in the multi-population analysis performed in the CelD GWAS by Dubois *et al*. (Dubois et al. 2010). Interestingly, different association patterns were observed for the two SNPs (Publication III, Table 1 and Table 2). Although neither of the SNPs was associated with CD or UC in general, rs9295089-C and rs1738074-T were associated with increased risk of several CD and/or UC subphenotypes, respectively. rs9295089-C was associated with increased risk (OR: 1.48-1.87) for
CD-subphenotypes, such as ileocolonic CD and extensive colitis, whereas rs1738074 was associated with protection (OR: 0.63) from UC in patients with a family history of UC. Neither of the SNPs was significant ($P_{EMP2} > 0.05$) after correcting for the number of markers used in the analysis. However, in the Finnish-Swedish UC meta-analysis, rs1738074-T was associated with protection (OR: 0.83) from UC and remained statistically significant ($P_{EMP2} < 0.05$) after correcting for the number of markers. Similarly, in the Finnish-Swedish UC subphenotype meta-analysis, rs1738074-T showed strong protective (OR: 0.61 and $P_{EMP2} < 0.005$) association with familial UC. These results suggest that rs1738074-T confers protection against familial UC. In the CelD GWAS (Dubois et al. 2010), this SNP was associated with increased risk for CelD. The same genetic variants thus appear to have opposing effects on risk for different diseases. A genetic variant (rs2451258-C) in the $TAGAP$ gene has been associated with increased risk for psoriasis (Tsoi et al. 2012). In contrast, rs1738074-T confers protection against multiple sclerosis (Patsopoulos et al. 2011) and T1D (Smyth et al. 2008). In addition other genetic variants in the $TAGAP$ gene have been associated with protection against rheumatoid arthritis (Eyre et al. 2010) and anal sepsis in ileocolonic CD (Connelly et al. 2012). Taken together, our results and previous studies suggest that the $TAGAP$ is a common susceptibility gene for several diseases. Furthermore, it is also highlighted that genes or genetic variants conferring susceptibility to one disease may be protective in another disease in a population-dependent manner.
Conclusions and future directions

In the first study of this thesis, the effect of gliadin peptides, the environmental trigger in CelD, on epithelial cell gene transcription was investigated. This study was undertaken as a result of previous studies suggesting direct effects of certain gliadin-derived peptides on enterocytes (Giovannini et al. 2003, Luciani et al. 2009, Sakly et al. 2006, Rivabene, Mancini & De Vincenzi 1999, Rivabene, Mancini & De Vincenzi 1999, Dolfini et al. 2002). In this study, a gene expression profile of Caco-2 cells exposed to PT-G and p31-43 peptide was obtained. Several genes were differentially regulated after stimulation with PT-G. However, they were also affected by the negative controls PT, PT-BSA and PT-L, suggesting that pepsin- and trypsin-derived peptides may affect, directly or indirectly, enterocyte gene transcription. This hypothesis gets support from previous studies (Johnston et al. 2012, Suen et al. 2010) as many genes differentially expressed after PT-G treatment are also differentially expressed after exposure to pepsin and trypsin. However, further studies on the role of pepsin and trypsin on enterocyte gene expression were beyond the scope of this thesis work and separate studies should be conducted to ascertain their role. In addition, the results suggest that gliadin may induce small differences in gene transcription. This may be particularly true for the p31-43 peptide as the microarray results could not be verified by qRT-PCR for any of the selected genes. The results from this study also suggest that most gliadin mediated effects in CelD may require the presence of other effector cells and chemical mediators. Based on the results obtained in this study, it is concluded that gliadin peptides alone have only small effects on enterocyte gene expression. The changes that gliadin peptides induce in gut epithelial cells of CelD patient are suggested to be a result of interactions between the immune system and gliadin peptides and it cannot be concluded that gliadin does not affect enterocyte gene transcription. An enterocyte gene expression study which also considers the possible effects of pepsin and trypsin and includes proper controls is thus highly recommended.

In the second study of this thesis, the association of rs601338-FUT2 with CelD, DH, CD, UC and IBD in the Finnish population was investigated. We found evidence for association of the rs601338-AA genotype with increased risk for CelD, but not with DH, most likely due to a small sample size. In contrast, no evidence for rs601338-A association with IBD was found. We conclude that the FUT2 non-secretor status (rs601338-AA) is associated with CelD but not with IBD in the Finnish population. The future course of research in this field should be to investigate how FUT2 defined gut microbiota affects the development of the immune system, subsequently affecting the risk of an individual to develop autoimmune diseases.
The third study focused on identifying which of the genetic variants that confer susceptibility to CelD also affect the risk of IBD in the Finnish and Swedish population.

Our results demonstrate that several CelD risk genes also confer susceptibility to IBD in the Finnish and the Swedish population. Some genes were associated with a particular disease subphenotype (e.g. familial UC, early onset UC/CD etc.) rather than with the disease in general (e.g. TAGAP, ELMO1 and UBE2L3). To conclude, several CelD risk loci were found to be associated with CD and/or UC or their associated subphenotypes in the Finnish and Swedish population. Furthermore, we show that the division of study subjects into different groups is useful for the identification of genes conferring susceptibility to a particular disease subphenotype, albeit it may lead to smaller sample sizes. To divide study subjects in subgroups based on their disease phenotype is an effective strategy for identifying phenotype-specific genes. However, the associated genes may not reach statistical significance after multiple testing corrections. Thus, future studies with a similar study design but larger datasets are recommended. If the sample sizes become large enough, one might consider more interaction studies. The future genetic studies may be combined with different sequencing methods (RNA sequencing, exome sequencing) to identify causal variants. Linkage studies in families could be combined with sequencing methods to identify rare disease variants. Different families may have different mutations in different genes, but those genes may participate in or affect common pathways.
Acknowledgements

This thesis work was carried out at the Department of Medical Genetics, University of Helsinki, Finland during the years 2006-2013. The financial support for this work was provided by the EU Commission Marie Curie Excellence Grant and the Academy of Finland. I want to express my gratitude to the funding bodies. Without their support, this research wouldn’t have been possible. All the study subjects are also thanked for their invaluable contribution to this study.

I would like to express my gratitude to Docent Päivi Saavalainen who introduced me to the world of genetics and immunological diseases. It has been a unique experience and I feel that the time spent here has changed me as a person and my view towards life. I have perhaps become more rational and certainly more patient. Had it not been so, I would have given up a long time ago. I am glad that despite a rather abnormally long period of setbacks, and that too in the beginning of my studies, I hung on and kept moving forward. Indeed “when the going gets tough, the tough get going”. Thank you Päivi for giving me the opportunity to work on this interesting subject!

Docent Elisabet Einarsdottir is also gratefully acknowledged for supervising me in the genetic projects. Elisabet and I joined the celiac disease research team on the same day and little did I know then that in a couple of years we will be doing association studies together. It has been a pleasure working with Elisabet and I thank her for being patient with me. This thesis could not have been possible without her help.

I want to thank all the coauthors for their contribution, in particular Prof. Markku Mäki’s celiac disease research group, Prof. Kimmo Kontula’s research group, Docent Mauro D’Amato and all the coauthors at Finnish Red Cross Blood Service for their contribution to the IBD and CelD projects.

Docent Kaija-Leena Kolho and Docent Hannu Torpeinen are acknowledged for reviewing the thesis. I believe that their suggestions have greatly enhanced the quality of this thesis and made it a worthy read.

I want to thank Docent Petteri Arstila and Docent Outi Monni for their comments and suggestions during the early stages of my doctoral studies.

I want to thank all former and present members of celiac disease research team, Emma Dukes, Hanne Ahola, Ovidiu Braicu, Fabiana Ziberna, Lotta Koskinen, Katri Haimila, Inna Sareneva, Tobias Freitag, Dawit Yohannes and Niina Wooley for their general company and for creating a pleasant working environment.
Prof. Juha Kere is acknowledged for all the support and stimulating discussions during group meetings. It has been a pleasure working in close collaboration with Prof. Kere and his research group in Helsinki. I want to thank all past and present members of the Kere research group for being wonderful coworkers.

I would like to express my gratitude to Prof. Markku Mäki and his research team for fruitful collaboration over the years. I have always admired Prof. Mäki for his passion for celiac disease research. It was indeed a pleasure to interact with him and he is a source of inspiration for many aspiring scientists.

My interest for research started during my M.Sc studies at Uppsala University and it started with a “by chance” meeting with Vinod Pant. Vinod is a great friend and although we talk only occasionally our friendship is just as cordial as it was years ago. I thank you Vinod for your friendship. I am very thankful to all my previous supervisors (Prof. Arne Östman, Prof. Tony Carr, Dr. James Brenton and Prof. Ian Hickson) who provided me with an opportunity to learn in their research groups. I met a lot of wonderful and talented people during these short, but very rewarding, training sessions. I am grateful that such an opportunity was bestowed upon me. I would also like to thank my friends at Biomedicum (Sofi & Siddesh Utge, Madhu & Biswajyoti Sahu, Krishna Chaithanya, Shashank Shekhar, Ashwini Yadav, Disha Malani and Ali Mahar) for their friendship and for soothing conversations during hard times.

This work (surprisingly!!) proved to be a mammoth task which could not have been possible without the support of my parents who believed in me. They have also patiently waited for this moment. This achievement belongs just as much to my wife Sherry who supported me in every possible way through thick and thin. Her inspirational and motivational quotes on Facebook kept me going. I am also lucky to have a stress buster at home, my son Ajit, who makes me forget everything at work and start over on a happy note. I want to dedicate this piece of work to my family.
References


Bisgaard, H., Li, N., Bonnelykke, K., Chawes, B.L., Skov, T., Paludan-Muller, G., Stokholm, J., Smith, B. & Krogfelt, K.A. 2011, "Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age", *The Journal of allergy and clinical immunology*.


Bodd, M., Raki, M., Tollefsen, S., Fallang, L.E., Bergseng, E., Lundin, K.E. & SOLlid, L.M. 2010, "HLA-DQ2-restricted gluten-reactive T cells produce IL-21 but not IL-17 or IL-22", *Mucosal immunology*, vol. 3, no. 6, pp. 594-601.


Finnish, Hungarian and Italian populations: association with IBD and psoriasis, and linkage to celiac disease, *BMC medical genetics*, vol. 10, pp. 8-2350-10-8.


Godkin, A., Frie, T., Davenport, M., Stevanovic, S., Willis, A., Jewell, D., Hill, A. & Rammensee, H.G. 1997, "Use of eluted peptide sequence data to identify the binding characteristics of peptides to the insulin-dependent diabetes susceptibility allele HLA-DQ8 (DQ 3.2)", International immunology, vol. 9, no. 6, pp. 905-911.


Halfvarson, J. 2011, "Genetics in twins with Crohn's disease: less pronounced than previously believed?", Inflammatory bowel diseases, vol. 17, no. 1, pp. 6-12.


Hidalgo, I.J., Raub, T.J. & Borchartd, R.T. 1989, "Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal epithelial permeability", *Gastroenterology*, vol. 96, no. 3, pp. 736-749.


Inagaki, H., Suzuki, T., Nomoto, K. & Yoshikai, Y. 1996, "Increased susceptibility to primary infection with Listeria monocytogenes in germfree mice may be due to lack of accumulation of L-selectin+ CD44+ T cells in sites of inflammation", Infection and immunity, vol. 64, no. 8, pp. 3280-3287.


King, C. & Sarvetnick, N. 2011, "The incidence of type-1 diabetes in NOD mice is modulated by restricted flora not germ-free conditions", *PloS one*, vol. 6, no. 2, pp. e17049.


Maiuri, M.C., De Stefano, D., Mele, G., Fecarotta, S., Greco, L., Troncone, R. & Carnuccio, R. 2003b, "Nuclear factor kappa B is activated in small intestinal mucosa of celiac patients", \textit{Journal of Molecular Medicine (Berlin, Germany)}, vol. 81, no. 6, pp. 373-379.


Mazzilli, M.C., Ferrante, P., Mariani, P., Martone, E., Petronzelli, F., Triglione, P. & Bonamico, M. 1992, "A study of Italian pediatric celiac disease patients confirms that the primary HLA association is to the DQ(alpha 1*0501, beta 1*0201) heterodimer", Human immunology, vol. 33, no. 2, pp. 133-139.


Moi, H. 1984, "Incidence and prevalence of dermatitis herpetiformis in a country in central Sweden, with comments on the course of the disease and IgA deposits as diagnostic criterion", Acta Dermato-Venereologica, vol. 64, no. 2, pp. 144-150.


chromosome 2q33 confers susceptibility to celiac disease in a way possibly distinct from that of type 1 diabetes and other chronic inflammatory disorders", *Tissue antigens*, vol. 56, no. 4, pp. 350-355.


Nilsen, E.M., Lundin, K.E., Krajci, P., Scott, H., Sollid, L.M. & Brandtzæg, P. 1995, "Gluten specific, HLA-DQ restricted T cells from coeliac mucosa produce cytokines with Th1 or Th0 profile dominated by interferon gamma", *Gut*, vol. 37, no. 6, pp. 766-776.


Ruiz-Palacios, G.M., Cervantes, L.E., Ramos, P., Chavez-Munguia, B. & Newburg, D.S. 2003, "Campylobacter jejuni binds intestinal H(O) antigen (Fuc alpha 1, 2Gal beta 1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection", *The Journal of biological chemistry*, vol. 278, no. 16, pp. 14112-14120.


Spurkland, A., Ingvarsson, G., Falk, E.S., Knutsen, I., Sollid, L.M. & Thorsby, E. 1997, "Dermatitis herpetiformis and celiac disease are both primarily associated with the HLA-DQ (alpha 1*0501, beta 1*02) or the HLA-DQ (alpha 1*03, beta 1*0302) heterodimers", Tissue antigens, vol. 49, no. 1, pp. 29-34.


Wang, D., Moothart, D.R., Lowy, D.R. & Qian, X. 2013, "The Expression of Glyceraldehyde-3-Phosphate Dehydrogenase Associated Cell Cycle (GACC) Genes Correlates with Cancer Stage and Poor Survival in Patients with Solid Tumors", *PloS one*, vol. 8, no. 4, pp. e61262.


Wellcome Trust Case Control Consortium 2007, "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls", *Nature*, vol. 447, no. 7145, pp. 661-678.


Williams, C.N., Kocher, K., Lander, E.S., Daly, M.J. & Rioux, J.D. 2002, "Using a genome-wide scan and meta-analysis to identify a novel IBD locus and confirm previously identified IBD loci", Inflammatory bowel diseases, vol. 8, no. 6, pp. 375-381.


